



## POLITECNICO DI TORINO Repository ISTITUZIONALE

### Cavity algorithms under global constraints: classical and quantum problems

*Original*

Cavity algorithms under global constraints: classical and quantum problems / Biazzo, Indaco. - (2014).

*Availability:*

This version is available at: 11583/2535118 since:

*Publisher:*

Politecnico di Torino

*Published*

DOI:10.6092/polito/porto/2535118

*Terms of use:*

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

POLITECNICO DI TORINO

DOCTORAL THESIS

---

**Cavity algorithms under global  
constraints: classical and  
quantum problems**

---

*Author:*

Indaco Biazzo

*Supervisor:*

Riccardo Zecchina

Statistical Physics and Interdisciplinary Applications

Dipartimento di scienza applicata e tecnologia

February 2014

*“..So is not true , as a recent article would have it, that we each should ”cultivate our own valley, and not attempt to build roads over the mountain ranges ... between the sciences.” Rather, we should recognize that such roads, while often the quickest shortcut to another part of our own science, are not visible from the viewpoint of one science alone.”* [\[1\]](#)

P. W. Anderson

# *Acknowledgements*

I wish to thank Valentina for everything, and Orlando for his smiles and for giving me the opportunity to sleep in the last year of my Ph.D. I am particularly grateful for freedom granted by Riccardo Zecchina and for the support given by Alfredo Braunstein. As well as I wish to acknowledge the lead, the help and the patience provided by Abolfazl Ramezanpour.

# Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>Contents</b>	<b>iii</b>
<b>Introduction</b>	<b>v</b>
<b>1 Cavity method and message passing algorithms</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Cavity Methods . . . . .	2
1.2.1 Marginal distribution . . . . .	5
1.2.2 Bethe free energy . . . . .	6
1.2.3 From tree to generic graphs . . . . .	7
1.3 Belief propagation . . . . .	8
1.3.1 Max-Sum . . . . .	9
1.3.2 BP equations on generic graph . . . . .	9
1.4 One-step replica symmetry breaking . . . . .	10
1.4.1 Sampling and stochastic optimizations . . . . .	13
<b>2 Optimization</b>	<b>15</b>
2.1 Introduction . . . . .	15
2.2 Combinatorial Optimization Problems . . . . .	17
2.2.1 Complexity scenario and NP class . . . . .	18
2.3 Prize collecting steiner tree (PCST) . . . . .	18
2.3.1 definition . . . . .	18
2.3.2 Rooted, depth bounded PCST and forests . . . . .	19
2.3.3 Local constraints . . . . .	19
2.4 Derivation of the message-passing cavity equations . . . . .	21
2.4.1 Max-sum: $\beta \rightarrow \infty$ limit . . . . .	22
2.4.2 Total fields . . . . .	23
2.4.3 Iterative dynamics and reinforcement . . . . .	24
2.4.4 Root choice . . . . .	25
2.5 other method . . . . .	25
2.5.1 integer linear programming . . . . .	26
2.5.2 Goemans-Williamson . . . . .	27
2.6 Computational Experiments . . . . .	27

2.6.1	Instances	27
2.6.2	Results	29
2.6.3	discussion	32
2.7	Post-processing and optimality	33
<b>3</b>	<b>Variational Quantum Cavity Method</b>	<b>35</b>
3.1	Introduction	35
3.2	Ground state approximate wave function	36
3.2.1	Quantum ising model	38
3.2.2	Pairwise model	39
3.2.3	Zero couplings: Mean field solution	41
3.2.4	Zero fields: Symmetric solution	41
3.2.5	General solution	42
3.2.6	Numerical results	43
3.2.6.1	Random coupling chain	43
3.2.6.2	Random graph	45
3.3	Low temperature excitations	45
3.3.1	Orthogonality constraints	48
3.3.2	The mean-field approximation	49
3.3.3	Beyond the mean-field approximation	52
3.3.4	Results	54
3.4	Discussion	56
<b>4</b>	<b>Variational Quantum Density Matrix</b>	<b>58</b>
4.1	Introduction	58
4.2	Imaginary time evolution	59
4.2.1	Imaginary time evolutions of a Bethe density matrix.	60
4.2.2	Tranverse ising model	62
4.2.3	Numerical results	64
4.3	Annealing of quantum systems	65
4.3.1	Annealing algorithm of quantum Ising model	68
4.3.2	Numerical result	70
<b>A</b>	<b>Post-processing and optimality proofs</b>	<b>72</b>
A.0.3	Proof of Theorem 2	73
<b>B</b>	<b>Locally and globally consistent reduced density matrices</b>	<b>75</b>
<b>C</b>	<b>Computing the reduced density matrices in the annealing algorithm</b>	<b>77</b>
	<b>Bibliography</b>	<b>79</b>

# Introduction

This thesis describes the path done by his author during his four years of PhD. The starting point of my PhD can be seen as relatively far from the final one. My hope is that at the end of the thesis the connections will appear.

The starting point was the study of optimization algorithms based on cavity method [2]. These algorithms have been developed to a high degree of complexity in the last decade and they are also known as message passing algorithms (MPAs).

MPAs are able to solve very difficult random combinatorial optimization problems (COP) [3–5] of central interest for the computer science community [6]. The main algorithmic approaches with which the MPAs have to compare are Integer lineal programming methods [7] and randomized searches.

My work has started by a question posed by my supervisor: what links can be found between those different approaches to the same problems? The starting aim of the PhD project was to explore the new ideas and algorithms that could result from a cross-fertilization between different approaches. During the first years we made a long and accurate comparison between different algorithms on a specific COP: the prize collecting Steiner tree problem. This was published in [8] and it is discussed in chapter 2. What has emerged from this analysis is the great performance of the cavity algorithm when compared with the integer linear programming techniques. The latter have a very long history, stated from 1951 [9], and their main ingredient consists in extending to real number the discrete variables of original problems.

What have these different approaches in common?

If we look at how such algorithms work at a very high level, we observe a sort of dynamical evolution of real quantities. In the case of message passing algorithms

we have messages, particular marginal probabilities defined over the discrete variables of the problem, changing all along the evolution of the algorithm. In the case of integer linear programming the relaxed discrete variables of the problem (to real values), change during running time. This is one of the main contact points between these different approaches which is discussed in more depth in chapter 2. Their temporal evolution is obviously very different from the Monte Carlo method where we have discrete configurations changing step by step.

Looking to MPAs as an evolution of probability distributions of discrete variables led me to find some possible links with many body quantum physics, where typically we deal with probability amplitudes over discrete variables. In the recent years several results have appeared concerning the extension of the cavity method and message passing technics to quantum context [10–14]. In 2012 Ramezanpour proposed a method for finding approximate ground state wave functions based on a new messages passing algorithm used in stochastic optimization[16]. Roughly speaking the idea was the following: giving a parametrization of a wave function with real parameters, the variational quantum problem of finding the optimal parameters maps onto a classical one which can be approached with the cavity method. Ramezanpour and I have extended this approach to find low excited states [17]. The results are presented in chapter 3.

What was not completely satisfying for me was the idea of working with two levels of probability distribution, one defined by the wave function and another defined over the parameters. The problem was solved using the imaginary time evolution operator which led to a great simplification of the problem and to a more accurate approximation of the optimal wave function. Moreover, this mapping allowed me to find a simple connection with an annealing in temperature of an approximate density matrix. These arguments are presented in chapter 4.



# Chapter 1

## Cavity method and message passing algorithms

### 1.1 Introduction

In this chapter I will describe the cavity method, a technique initially invented to deal with the Sherrington Kirkpatrick model of spin glasses, alternative to the replica approach [18]. The cavity method uses a probabilistic approach that allows to have more intuitive feeling on what is going on compared to the replica formalism [2]. This method was initially applied on infinite dimension systems, then in 2001 it was extended to finite temperature systems with a finite number of neighbors like Bethe lattice and random graphs [19].

The interest of such systems relies on several different aspects. On one hand we may hope to get a better knowledge on finite dimension problems due to notions of neighborhood included by these systems. On the other hand we can have the possibility to solve these problems with iterative methods. The cavity method, in fact, is a generalization of the Bethe-Peierls method [20]. Another reason comes from the strong connection with optimization problems which typically have finite connectivity structures.

The cavity method was also extended to deal with zero temperature systems; this was a crucial step in order to develop new class of algorithms that allow to solve difficult optimization problems [3–5]. The cavity method has several levels

of approximation, from the replica symmetric case to replica symmetric breaking solutions [2].

The cavity method can be applied to a single instance of a problem under the form of a message passing algorithm. The first level of approximation, the replica symmetric case, is equal to the so called Belief propagation (BP) equations. The BP equations have been rediscovered many times in many different contexts. In Bayesian inference context the BP equations were developed by Pearl [21], in decoding context by Gallager [22]. In 1935, in physics, Bethe used them on a homogenous system, and their generalization to inhomogeneous systems waited until the application of Bethe's method to spin glasses [23]. An algorithm that utilizes the 1RSB cavity method to find solutions for a single instance of a problem was done by Mézard et al. [3, 4]. The idea that the 1RSB method can be derived applying the BP equations over an auxiliary model, solved by a low layer BP equations, appears for the first time in the articles [24, 25]. The idea of more level of BP equations has led to write new algorithms that can solve, in efficient ways, very difficult optimization problems, like stochastic optimization and sampling over a huge space of parameters that should minimize an energy function [15, 16, 26].

In the section 1.2 I will show the cavity method at replica symmetric level, then in the sec 1.3 are presented the BP and the Max-Sum equations and their generalization to generic graph. In the last section 1.4 the 1RSB equations are derived, and it is explained the procedure to solve stochastic optimization problems with several layers of BP equations.

## 1.2 Cavity Methods

Statistical physics, in general, provides a framework for relating the microscopic properties of individual objects to the macroscopic or bulk properties of the system under study. The statistical physics started, at the end of XIX century, on purpose to explain thermodynamics as a natural result of statistics, classical mechanics, and then quantum mechanics at the microscopic level. Central in the statistical physics, at least in equilibrium statistical physics, is the partition function  $Z$ . Given all possible configuration  $\mathbf{q}$  of a system, the  $Z$  is defined as:

$$Z = \sum_{\mathbf{q}} e^{-\beta H(\mathbf{q})}, \quad (1.1)$$

where  $\beta = 1/T$  is the inverse of temperature, and  $H(\mathbf{q})$  is a function depending on the configuration of the system, typically the energy function. The partition function is directly connected to a macroscopic quantity, the Helmholtz free energy:

$$F = -\frac{1}{\beta} \ln Z \quad (1.2)$$

where the Helmholtz free energy is the average energy of the system minus the temperature times the entropy,  $F = \langle H \rangle - \frac{1}{\beta} S$ .

Given the partition function is also possible to define the probability of a configuration:

$$P(\mathbf{q}, \beta) = \frac{e^{-\beta E(\mathbf{q})}}{Z}. \quad (1.3)$$

Given this probability measure it is possible to compute explicitly macroscopic quantity e.g., the average energy

$$E = \sum_{\mathbf{q}} P(\mathbf{q}, \beta) H(\mathbf{q}), \quad (1.4)$$

the entropy

$$S = -\sum_{\mathbf{q}} P(\mathbf{q}, \beta) \ln P(\mathbf{q}, \beta). \quad (1.5)$$

The number of configurations typically grows exponentially with the number of elements that belong to the system. This makes, in general, a hard task computing the partition function.

This thesis is devoted to a new class of algorithms that permits to compute, in general, an approximate probability measure of systems in efficient way. The theoretical substrate of these algorithms is the cavity method described in the next section. The cavity method, initially invented to deal with the Sherrington Kirkpatrick model of spin glasses, is a powerful method to compute the properties of systems in many condensed matter and optimization problems [2]. There are several levels of approximation, in this section I will present only the first one, called replica symmetric solution, for further informations see [27, 28].

Before explaining the cavity method, I will show how to solve classical statistical physics models on trees. Consider a system with  $N$  discrete variables  $\sigma$  that can take only two values  $\{-1, 1\}$ . For simplicity, the system has only a two body

interaction:

$$E(\boldsymbol{\sigma}) = \sum_{\langle i,j \rangle} E_{ij}(\sigma_i \sigma_j) \quad (1.6)$$

where the  $\langle i, j \rangle$  represent the links present in the system. The links generate a graph and we consider now a tree structure. In this case the computation of the partition function can be achieved using the recursion properties of the tree. I define  $\partial i$  as the set of variables adjacent to a given vertex  $i$ , and  $\partial i \setminus j$  as vertices around  $i$  distinct from  $j$ . Consider the quantity  $Z_{i \rightarrow j}(\sigma_i)$ , for two adjacent sites  $i$  and  $j$ , as the partial partition function for the subtree rooted at  $i$ , excluding the branch directed towards  $j$ , with a fixed value of  $\sigma_i$ . Then consider the quantity  $Z_i(\sigma_i)$  as the partition function of the whole tree with a fixed value of  $\sigma_i$ . These quantities can be computed according to recursion rules:

$$Z_{i \rightarrow j}(\sigma_j) = \sum_{\sigma_i} \psi_{ik}(\sigma_i, \sigma_k) \prod_{k \in \partial i \setminus j} Z_{k \rightarrow i}(\sigma_i), \quad (1.7)$$

$$Z_i(\sigma_i) = \prod_{j \in \partial i} Z_{j \rightarrow i}(\sigma_i), \quad (1.8)$$

where  $\psi_{ij} = e^{-\beta E_{ij}}$ . In general it's useful to work with normalized quantities which can be interpreted as probability laws for the variables. The first one is the marginal probability law of the variable  $\sigma_j$  in a modified system where all the links around  $j$  have been removed but  $\langle i, j \rangle$ :

$$\mu_{i \rightarrow j}(\sigma_j) = \frac{Z_{i \rightarrow j}(\sigma_j)}{\sum_{\sigma_j} Z_{i \rightarrow j}(\sigma_j)} = \frac{1}{z_{i \rightarrow j}} \sum_{\sigma_i} \psi_{ik}(\sigma_i, \sigma_k) \prod_{k \in \partial i \setminus j} \mu_{k \rightarrow i}(\sigma_i). \quad (1.9)$$

The second one is the marginal of  $\sigma_i$  on the whole system:

$$\rho_i(\sigma_i) = \frac{Z_i(\sigma_i)}{\sum_{\sigma_i} Z_i(\sigma_i)} = \frac{1}{z_i} \prod_{j \in \partial i} \mu_{j \rightarrow i}(\sigma_i) \quad (1.10)$$

The quantities  $z_{i \rightarrow j}$  and  $z_i$  are normalization constants. If we consider a generic system composed by  $N$  variables  $\{\boldsymbol{\sigma}\}$  the computation of the partition function (1.1) involves a sum over  $2^N$  configurations. But if we are on tree, and if we start to iterate the (1.7) from leaves, **the computation of the partition function scales linearly with the system size  $N$ .**

### 1.2.1 Marginal distribution

The marginal distributions of a set of connected variables on tree admit an explicit expression in terms of messages. Let  $\sigma_R$  be a subset of variables of the system, let  $E_R$  be the subset of energy terms that connect or belong to variables that are in the set  $\sigma_R$ , let  $\partial R$  be the subset of variables that are connected to the variables  $\sigma_R$  but don't belong to it. It's easy to show that the marginal distribution is:

$$\rho(\sigma_R) = \frac{1}{Z_R} \prod_{\langle a \rangle \in E_R} \psi_a(\sigma_a) \prod_{i \in \sigma_R} \prod_{j \in \partial i \cap \partial R} \mu_{j \rightarrow i}(\sigma_i). \quad (1.11)$$

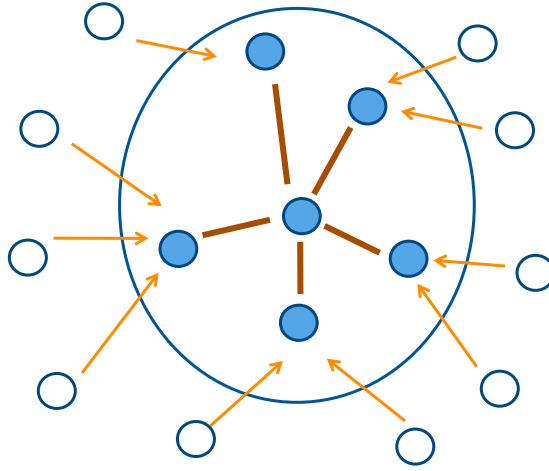


FIGURE 1.1: A graphical picture of the computation of a marginal region. The variables that belong to the region of the marginal probability are the blue filled circles. White filled circles are the variables in the neighborhood of the region  $R$ . Red lines represent the energy terms that are inside the region  $R$  and orange arrows are the messages entering in the region  $R$ .

The figure 1.1 gives a graphical picture of the equations. For instance the marginal probability of two adjacent variables is:

$$\rho_{i,j}(\sigma_i, \sigma_j) = \frac{1}{z_{ij}} \psi_{ij}(\sigma_i, \sigma_j) \prod_{p \in \partial i \setminus j} \mu_{p \rightarrow i}(\sigma_i) \prod_{q \in \partial j \setminus i} \mu_{q \rightarrow j}(\sigma_j) \quad (1.12)$$

where  $z_{ij} = \sum_{\sigma_i, \sigma_j} \psi_{ij} \prod_{p \in \partial i \setminus j} \mu_{p \rightarrow i}(\sigma_i) \prod_{q \in \partial j \setminus i} \mu_{q \rightarrow j}(\sigma_j)$  is the normalization.

### 1.2.2 Bethe free energy

When the structure of the system under study is a tree, it's possible to factorize the probability distribution by a product of local marginals. The explicitly form is:

$$P(\boldsymbol{\sigma}) = \frac{\prod_{\langle i,j \rangle} e^{\beta E_{ij}}}{Z} = \prod_i \rho_i \prod_{\langle i,j \rangle} \frac{\rho_{ij}}{\rho_i \rho_j} \quad (1.13)$$

In order to demonstrate the equation above, we have to consider that:

1.

$$\sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}) = 1 \quad (1.14)$$

$$\sum_{\boldsymbol{\sigma}} \prod_i \rho_i \prod_{\langle i,j \rangle} \frac{\rho_{ij}}{\rho_i \rho_j} = 1. \quad (1.15)$$

The equation (1.14) is true by definition. Considering the marginalization constraints  $\sum_{\sigma_j} \rho_{i,j}(\sigma_i, \sigma_j) = \rho_i(\sigma_i)$ , it is possible to recover the eq.(1.15) just starting the sum over all configurations from leaves: every factor iteratively simplifies.

2.

$$\prod_i \rho_i \prod_{\langle i,j \rangle} \frac{\rho_{ij}}{\rho_i \rho_j} = \quad (1.16)$$

$$= \prod_i \frac{\prod_{j \in \partial i} \mu_{j \rightarrow i}(\sigma_i)}{z_i} \prod_{i,j} \frac{z_i z_j \psi_{ij} \prod_{p \in \partial i \setminus j} \mu_{p \rightarrow i}(\sigma_i) \prod_{q \in \partial j \setminus i} \mu_{q \rightarrow j}(\sigma_j)}{z_{i,j} \prod_{p \in \partial i} \mu_{p \rightarrow i}(\sigma_i) \prod_{q \in \partial j} \mu_{q \rightarrow j}(\sigma_j)} \quad (1.17)$$

$$= \prod_i \frac{1}{z_i} \prod_{i,j} \frac{z_i z_j \psi_{ij}}{z_{i,j}} \quad (1.18)$$

$$\propto \prod_{i,j} \psi_{ij} = \prod_{\langle i,j \rangle} e^{\beta E_{ij}} \quad (1.19)$$

$$\propto P(\boldsymbol{\sigma}), \quad (1.20)$$

where to obtained the final outcome I just substituted the definition of the marginals eq.(1.10) and eq.(1.12), and simplified the remaining terms.

The point one (both quantities sum to one) and the point two (they are proportional each other) imply that the two quantities are the same and the eq. (1.13) is correct. Moreover it's easy to show that:

$$Z = \prod_i z_i \prod_{\langle i,j \rangle} \frac{z_{ij}}{z_i z_j} \quad (1.21)$$

Given the eq.(1.13) for the probability distribution and the eq.(1.21), the entropy reads:

$$S = - \sum_{\substack{i,j \\ \sigma_i \sigma_j}} \rho_{ij} \ln \rho_{ij} + (d_i - 1) \sum_{i, \sigma_i} \rho_i \ln \rho_i, \quad (1.22)$$

where  $d_i$  is the number of adjacent sites of  $i$ . The average energy is:

$$E = \sum_{\substack{i,j \\ \sigma_i \sigma_j}} E_{ij} \rho_{ij}. \quad (1.23)$$

Then, when we are dealing with tree structures also the free energy takes a simpler form, called Bethe free energy:

$$F = E - \frac{1}{\beta} S = -\frac{1}{\beta} \left[ \sum_{\substack{i,j \\ \sigma_i \sigma_j}} \rho_{ij} \ln \frac{\rho_{ij}}{\psi_{ij}} - (d_i - 1) \sum_{i, \sigma_i} \rho_i \ln \rho_i \right] \quad (1.24)$$

### 1.2.3 From tree to generic graphs

Now we suppose that the interaction graph is no longer a tree. We Consider however a site  $i$  where we cut all the edges around it but one  $\langle i, j \rangle$ . We make a "cavity". Then we want to know what is the marginal probability law  $\mu_{j \rightarrow i}(\sigma_i)$  of the variable  $i$  in that modified graph. It is:

$$\mu_{j \rightarrow i}(\sigma_i) = \frac{1}{z_{i \rightarrow j}} \sum_{\sigma_j} \psi_{ij} \mu_{\partial j \setminus i \rightarrow j}(\sigma_j) \quad (1.25)$$

where we introduced a multi-variable cavity field  $\mu_{\partial j \setminus i \rightarrow j}(\sigma_j)$  that describes the probability distribution of  $\sigma_j$  due to the presence of the link connection between  $j$  and the variables  $\partial j \setminus i$ . The eq.(1.25) is not in a closed form, if we want to close

it we have to factorize the  $\mu_{\partial j \setminus i \rightarrow j}$ :

$$\mu_{\partial j \setminus i \rightarrow j}(\sigma_j) = \prod_{k \in \partial j \setminus i} \mu_{k \rightarrow j}(\sigma_j). \quad (1.26)$$

In generic graphs the assumption (1.26) is not correct, but it can be a good or bad assumption depending on problems. **The (replica symmetric) cavity method consists in using recursion equations (1.9), assuming the (1.26), to compute the free energy and other quantities of interest of the system.** The recursive equation (1.9) is at the base of the message passing algorithms used in this thesis.

### 1.3 Belief propagation

The cavity method described in the section 1.2 has a natural algorithmic interpretation. On a generic graph, with  $\|E\|$  number of edges, we have  $2\|E\|$  cavity equations (1.9) and  $2\|E\|$  number of unknown messages  $\mu_{i \rightarrow j}$ . We can hope for unique solution and try to solve that system of equations (1.9) by recursion:

$$\mu_{i \rightarrow j}^{t+1}(\sigma_j) = \frac{1}{z_{i \rightarrow j}} \sum_{\sigma_i} \psi_{ik}(\sigma_i, \sigma_k) \prod_{k \in \partial i \setminus j} \mu_{k \rightarrow i}^t(\sigma_i). \quad (1.27)$$

This is known as belief propagation equations (BP). When such equations hope-fully converge we have a set of messages  $\{\mu_{i \rightarrow j}^*\}$  through which we can compute quantities of interest like one-body marginal:

$$\rho_i = \frac{1}{z_i} \prod_{j \in \partial i} \mu_{j \rightarrow i}^*(\sigma_i). \quad (1.28)$$

When we are on tree the quantities that we compute by the messages are exact, but on generic graph the quantities computed are approximate. Typically the belief propagation equations (1.27) give very precise result on sparse random graphs. They are locally tree like and the length of loop is  $\sim \log N$  where  $N$  is the number of sites or variables of the system.

The usual procedure to find the fixed point of the eq.(1.27) is to start from random values of messages and to update in random order the messages on each edge.



### 1.3.1 Max-Sum

When we are interested in ground-state configurations it is possible to rewrite the BP equations directly at zero temperature. The probability distribution of the system  $P(\boldsymbol{\sigma}) = \frac{e^{\beta H}}{Z}$ , when  $\beta \rightarrow \infty$ , is concentrated over the ground-state configurations. Taking the limit  $\beta \rightarrow \infty$  for the BP equations read:

$$\lim_{\beta \rightarrow \infty} \mu_{i \rightarrow j}^{t+1}(\sigma_j) = \lim_{\beta \rightarrow \infty} \left( \frac{1}{z_{i \rightarrow j}} \sum_{\sigma_i} e^{-\beta E_{ij}}(\sigma_i, \sigma_k) \prod_{k \in \partial i \setminus j} \mu_{k \rightarrow i}^t(\sigma_i) \right) \quad (1.29)$$

$$= \max_{\sigma_i} \left[ \lim_{\beta \rightarrow \infty} \left( \frac{1}{z_{i \rightarrow j}} e^{-\beta E_{ij}}(\sigma_i, \sigma_k) \prod_{k \in \partial i \setminus j} \mu_{k \rightarrow i}^t(\sigma_i) \right) \right]. \quad (1.30)$$

Defining  $M_{i \rightarrow j} = \frac{1}{\beta} \log \mu_{i \rightarrow j}$  we can write:

$$M_{i \rightarrow j}^{t+1}(\sigma_j) = \max_{\sigma_i} \left[ -E_{ij} + \sum_{k \in \partial i \setminus j} M_{k \rightarrow i}^t(\sigma_i) \right] + C_{i \rightarrow j}. \quad (1.31)$$

These are called Max-Sum equations. The constant  $C_{i \rightarrow j}$  is typically set imposing that  $\max_{\sigma_j} M_{i \rightarrow j} = 0$ . When the equations, hopefully, converge we can find the configuration that minimizes the energy by:

$$\sigma_i^* = \arg \max_{\sigma_i} \left\{ -E_{ij} + \sum_{j \in \partial i} M_{j \rightarrow i}^* \right\} \quad (1.32)$$

### 1.3.2 BP equations on generic graph

It's easy to generalize the BP and Max-Sum equations to a generic interaction graph. Consider a system that has  $M$  interactions  $\boldsymbol{\psi}$  and  $N$  variables  $\boldsymbol{\sigma}$ . We define the variables  $\boldsymbol{\sigma}_{\partial a}$  as the neighborhood of the interaction  $\psi_a(\boldsymbol{\sigma}_{\partial a})$ , they are the variables involved in the interaction  $a$ . Then we define the interaction  $a \in \partial i$  neighborhood of a variable  $i$  as the interaction  $a$  that includes  $i$ . Giving a general probability distribution:

$$P(\boldsymbol{\sigma}) = \frac{\prod_a \psi_a(\boldsymbol{\sigma}_{\partial a})}{Z} \quad (1.33)$$

where  $Z$  is the usual normalization constant, we have two sets of BP equations:

$$\mu_{i \rightarrow a}^{t+1}(\sigma_i) = \frac{1}{z_{i \rightarrow a}} \prod_{b \in \partial i \setminus a} \mu_{a \rightarrow i}^t(\sigma_i) \quad (1.34)$$

$$\mu_{a \rightarrow i}^{t+1}(\sigma_i) = \frac{1}{z_{a \rightarrow i}} \sum_{\mathbf{x}_{\partial a \setminus i}} \psi_a(\boldsymbol{\sigma}_{\partial a}) \prod_{k \in \partial a \setminus i} \mu_{k \rightarrow a}^t(\sigma_k). \quad (1.35)$$

These sets of equation permit to compute in a efficient way all the quantities of interest of a system, in an exact way if the structure is a tree, in an approximate way otherwise. The Max-Sum equations can be derived in the same way showed in the sec.(1.3.1). The free energy reads:

$$F = \sum_a F_a + \sum_i F_i - \sum_{ia} F_{ia} \quad (1.36)$$

where

$$F_a = \log \left[ \sum_{\boldsymbol{\sigma}_{\partial a}} \psi_a(\boldsymbol{\sigma}_{\partial a}) \prod_{i \in \partial a} \mu_{i \rightarrow a}(\sigma_i) \right], \quad (1.37)$$

$$F_i = \log \left[ \sum_{\sigma_i} \prod_{b \in \partial i} \mu_{b \rightarrow i}(\sigma_i) \right] \quad (1.38)$$

$$F_{ai} = \log \left[ \sum_{\sigma_i} \rho_{i \rightarrow a}(\sigma_i) \mu_{b \rightarrow i}(\sigma_i) \right] \quad (1.39)$$

## 1.4 One-step replica symmetry breaking

The effectiveness of belief propagation depends on one basic assumption: when a node is pruned from the system, the adjacent variables become weakly correlated with respect to the resulting distribution. This hypothesis may break down either because of the existence of small loops in the factor graph or because variables are correlated at large distances. When we consider locally tree like systems the grow of the long distances correlation is responsible for the failure of BP equations.

The one-step replica symmetry breaking (1RSB) cavity method is a non rigorous approach to deal with this phenomenon. The long range correlation can be made in connection to the appearance of metastable states. The metastable states can be described by pure states. It is a general result of statistical physics [2, 29] that it is always possible to decompose the equilibrium probability distribution as a

sum over pure states. In principle it is possible to select one pure state, among many, by adding an external field to the system: the probability distribution of pure states is the limit to zero field of the Gibbs measure of the system. The overall probability measure is the sum over the distribution of each pure state:

$$P(\boldsymbol{\sigma}) = \frac{e^{-\beta H}}{Z} = \sum_{\alpha} \omega_{\alpha} P^{\alpha}(\boldsymbol{\sigma}), \quad (1.40)$$

where the  $\omega_{\alpha}$  are positive weights normalized to one. An important property of pure states is the cluster property, i.e. the correlation vanishes at large distances [30]. One of the basic assumption of 1RSB is on the values of  $\omega_{\alpha}$ , they are proportional to the Bethe free energy of the pure state:

$$\omega_{\alpha} = \frac{e^{F_{\alpha}}}{\Xi}, \quad \Xi = \sum_{\alpha} \omega_{\alpha}. \quad (1.41)$$

The second basic assumption is that the number of Bethe pure state is equal to the number of solutions of BP equations (1.9). We introduce an auxiliary statistical physics problem through the definition of a canonical distribution over this Bethe pure state. We assign to each pure state the probability  $\omega_{\alpha} = \frac{e^{xF_{\alpha}}}{\Xi}$  where  $x$  play the role of the inverse temperature, called the Parisi 1RSB parameter. So we have the partition function:

$$\Xi(x) = \sum_{\alpha} e^{xF_{\alpha}}. \quad (1.42)$$

The 1RSB can be summarized as: *Introducing a Boltzmann measure over Bethe measure, and solve it using BP equations.* We now deal with two systems, two levels, the first one is the usual one, and the second one has the messages over the first system as variables. We consider now the simpler version of BP equations described in the section 1.2, where we consider only pairwise interactions:

$$P(\boldsymbol{\sigma}) = \frac{\prod_{ij} \psi_{ij}(\sigma_i, \sigma_j)}{Z}. \quad (1.43)$$

In order to compute the number of solutions of BP equations, we rewrite them:

$$\mu_{i \rightarrow j}(\sigma_j) = \frac{1}{Z_{i \rightarrow j}} \sum_{\sigma_i} \psi_{ik}(\sigma_i, \sigma_k) \prod_{k \in \partial i \setminus j} \mu_{k \rightarrow i}(\sigma_i) = f_{i \rightarrow j}[\mu_{k \rightarrow i}, k \in \partial i \setminus j]. \quad (1.44)$$

The Bethe free energy (1.2.2) of each state is:

$$-\beta F_\alpha [\boldsymbol{\mu}_{i \rightarrow j}^\alpha] = \sum_{i,j} \log \frac{z_{ij}^\alpha}{z_i^\alpha z_j^\alpha} + \sum_i \log z_i^\alpha \quad (1.45)$$

where:

$$z_{ij}^\alpha = \sum_{\sigma_i \sigma_j} \psi_{ij} \prod_{p \in \partial i \setminus j} \mu_{p \rightarrow i}^\alpha(\sigma_i) \prod_{q \in \partial j \setminus i} \mu_{q \rightarrow j}^\alpha(\sigma_j) \quad (1.46)$$

$$z_i^\alpha = \sum_{\sigma_i} \prod_{j \in \partial i} \mu_{j \rightarrow i}^\alpha(\sigma_i). \quad (1.47)$$

The set of messages  $\boldsymbol{\mu}_{i \rightarrow j}^\alpha$  is one solution of the equations (1.44). Now it is possible to compute the partition function:

$$\Xi = \sum_{\alpha} e^{x F_\alpha} \quad (1.48)$$

$$= \int \mathcal{D} \boldsymbol{\mu}_{i \rightarrow j} \sum_{\alpha} e^{x F_\alpha [\boldsymbol{\mu}_{i \rightarrow j}]} \prod_{ij} \delta(\mu_{i \rightarrow j} - f_{i \rightarrow j}) \delta(\mu_{j \rightarrow i} - f_{j \rightarrow i}) \quad (1.49)$$

$$= \int \mathcal{D} \boldsymbol{\mu}_{i \rightarrow j} \prod_i \left[ z_i \prod_{j \in \partial i} \delta(\mu_{j \rightarrow i} - f_{j \rightarrow i}) \right] \prod_{ij} \frac{z_{ij}}{z_i z_j} \quad (1.50)$$

We solve this problem using BP equations:

$$\begin{aligned} Q_{i \rightarrow j}(\mu_{i \rightarrow j}, \mu_{j \rightarrow i}) &= \\ &= \frac{1}{\mathcal{Z}_{i \rightarrow j}} \sum_{\substack{\{\mu_{k \rightarrow i}, \mu_{i \rightarrow k}\} \\ k \in \partial i \setminus j}} z_{i \rightarrow j} \prod_{q \in \partial i} \delta(\mu_{i \rightarrow q} - f_{i \rightarrow q}) \prod_{k \in \partial i \setminus j} Q_{k \rightarrow i}(\mu_{k \rightarrow i}, \mu_{i \rightarrow k}), \end{aligned} \quad (1.51)$$

where we use the identity  $z_{ij} = z_j z_{i \rightarrow j}$ . The message equation can be further simplified summing over the  $\mu_{k \rightarrow i}, k \in \partial i \setminus j$  and using the delta function. What we obtain is the following equation:

$$Q_{i \rightarrow j}(\mu_{j \rightarrow i}) = \frac{1}{\mathcal{Z}_{i \rightarrow j}} \sum_{\{\mu_{i \rightarrow k}\}_{k \in \partial i \setminus j}} z_{i \rightarrow j} \delta(\mu_{i \rightarrow j} - f_{i \rightarrow j}) \prod_{k \in \partial i \setminus j} Q_{k \rightarrow i}(\mu_{k \rightarrow i}). \quad (1.52)$$

The set of messages  $\{\boldsymbol{Q}_{i \rightarrow j}\}$  allows to compute the partition function:

$$\Xi = \prod_i z_i \prod_{ij} \frac{z_{ij}}{z_i z_j} \quad (1.53)$$

where:

$$\mathcal{Z}_{ij} = \sum_{\mu_{i \rightarrow j}, \mu_{j \rightarrow i}} z_{ij} Q_{j \rightarrow i} Q_{i \rightarrow j} \quad (1.54)$$

$$\mathcal{Z}_i = \sum_{\mu_{\{j \rightarrow i\}}, j \in \partial i} z_i \prod_{j \in \partial i} Q_{j \rightarrow i}. \quad (1.55)$$

The 1RSB equations have been used to devise new classes of algorithms. One example is the Survey Propagation algorithm that is very effective to solve hard optimization problems like K-SAT [4].

### 1.4.1 Sampling and stochastic optimizations

Most real-world optimization problems involve uncertainty: the precise value of some of the parameters is often unknown, either because they are measured with insufficient accuracy, or because they are stochastic in nature and revealed in subsequent stages. The objective of the optimization process is thus to find solutions which are optimal in some probabilistic sense, a fact which introduces fundamental conceptual and computational challenges [31]. Examples of stochastic optimization problems can be found in many areas of sciences ranging from resource allocation and robust design problems in economics and engineering, to problems in physics and biology. A key computational difficulty in optimization under uncertainty comes from the size of the uncertainty space, which is often huge and leads to very large-scale optimization models, e.g. the sampling. Moreover, optimizing under uncertainty is further complicated by the discrete nature of decision variables. The idea is to solve such problems using equations that are similar to the 1RSB method [16]. We consider now for simplicity a particular stochastic optimization problem, the two stage problem. The input of the problem is an energy function  $E(\sigma_1, \sigma_2, \mathbf{t})$ . We have two sets of discrete variables  $(\sigma_1, \sigma_2)$  and an independent set of stochastic parameters  $\mathbf{t}$ , extracted from a distribution  $P(\mathbf{t})$ . The goal is to find  $\sigma_1$  that minimize the average energy obtained with the following process: first  $\sigma_1$  are fixed, then  $\mathbf{t}$  are extracted, and finally we minimize over the  $\sigma_2$  variables. The problem can be expressed as follow:

$$\sigma_1^* = \arg \min_{\sigma_1} \int d\mathbf{t} P(\mathbf{t}) \min_{\sigma_2} E(\sigma_1, \sigma_2, \mathbf{t}) \quad (1.56)$$

The method, that [Altarelli et al.](#) proposed, consists in performing the minimizations and expectations in the previous equations by means of respectively Max-Sum and BP equations. The key point of this method is that in both cases the optimal or expected value of the cost function can be computed in terms of the messages, and these are again a sum of local terms. This allows to perform also the next minimization or expectation operation by Max-Sum or BP equations, by considering the messages of the previous problem as the variables of a new one. I do not give here a complete derivation of equations that can be found in the article [\[26\]](#). The procedure can be summarized as:

1. First we write the Max-Sum(MS) equations for the inner minimum of eq. [\(1.4.1\)](#); we have a set of messages  $\mathbf{m}$ . We build the distribution  $\mathcal{Q}(\mathbf{t}, \mathbf{m}) \propto P(\mathbf{t}) \mathbb{1}_{MS}(\mathbf{m}, \mathbf{t})$ , where  $\mathbb{1}_{MS}(\mathbf{m}, \mathbf{t})$  is the Max-Sum equation constraint over the values of messages  $\mathbf{m}$  like in 1RSB equations [\(1.4\)](#).
2. Write the BP equations for  $\mathcal{Q}$ , with messages  $\mathbf{Q}$ .
3. Consider now the average energy:

$$E(\boldsymbol{\sigma}_1, \mathbf{Q}) = \langle E(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \mathbf{t}) \rangle_{\mathbf{Q}}. \quad (1.57)$$

It depends by  $\boldsymbol{\sigma}_1$ , that we have considered until now constant, and by BP messages  $\mathbf{Q}$ .

4. At the end we perform the last minimization with Max-Sum equations of  $E(\boldsymbol{\sigma}_1, \mathbf{Q})$  over both  $\mathbf{Q}$  and  $\boldsymbol{\sigma}_1$  imposing the BP constraints over the messages  $\mathbf{Q}$ .

This algorithm is able to find better solutions than the standard algorithms (greedy approach, linear programming) for a large set of hard instances [\[16, 26\]](#).

# Chapter 2

## Optimization

### 2.1 Introduction

In the last two decades there have been a huge convergence of interests between different fields of science: coding theory, statistical mechanics, statistical inference and computer science. Probabilistic models and technique have been the main underline reason for this convergence. The probabilistic approach in statistical physics [32], statistical inference [33] and in information theory [21, 34] has a long history. Technics and concepts borrowed from statistical physics have led to new insights into theoretical computer science, see for instance [35, 36].

The connection between these fields is not only formal, due to the common probabilistic background, but they share problems, questions and results of central interest for each of them. This set of problems and technics can be named as “theory of large graphical model” [37]. The typical definition is the following: “a large set of random variables taking values in a finite (typically quite small) alphabet with a *local* dependency structure; this local dependency structure is conveniently described by an appropriate graph.” [37]. In computer science, combinatorial optimization problems (COPs) plays a central role. COPs typically consist of finding configurations of a system that minimize a cost function. Many COPs belong to the NP-complete class. It means, roughly speaking, that the computer time to find a problem’s solution, in the worst case, grows exponentially with the system size. In physics, optimization problems are equal to find the ground state energy of a system. Physics models, like the spin glass, have been proven belong to the NP-complete class. Physicists, usually, are interested in the typical solution instead

of a worst case analysis [38]. In these last decades the development of a typical case analysis has become a major goal [39–41], motivated also by experimental evidence that many instances of NP-complete problems are easy to solve [36]. In recent years the cavity method, described in the first chapter, has led to design a family of algorithmic techniques for combinatorial optimization problems. In spite of the numerical evidence of the large potentiality of these techniques in terms of efficiency and quality of results, their use in real-world problems has still to be fully expressed. The main reasons for this reside in the fact that the derivation of the equations underlying the algorithms are in many cases non-trivial and that the rigorous and numerical analyses of the cavity equations are still largely incomplete. Both rigorous results and benchmarking would play an important role in helping the process of integrating message passing algorithms (MPAs) with the existing techniques.

In what follows we focus on a very well known NP-hard optimization problem over networks, the so-called Prize Collecting Steiner Tree problem on graphs (PCST). The PCST problem can be stated in general terms as the problem of finding a connected subgraph of minimum cost. It has applications in many areas ranging from biology, e.g. finding protein associations in cell signaling [42–44], to network technologies, e.g. finding optimal ways to deploy fiber optic and heating networks for households and industries [45].

In this chapter it's shown how `msgsteiner` – an algorithm derived from the zero temperature cavity equations for the problem of inferring protein associations in cell signaling [43, 44] – compares with state-of-the-art techniques on benchmarks problem instances. Specifically, we provide comparison results with an enhanced derivative of the Goemans-Williamson heuristics (MGW) [46, 47] and with the DHEA solver [48], a Branch and Cut Linear/Integer Programming based approach. We made the comparison both on random networks and in known benchmarks. We show that `msgsteiner` typically outperforms the state-of-the-art algorithms in the largest instances of the PCST problem both in the values of the optimum and in running time.

Finally, it's shown how some aspects of the solutions can be provably characterized. Specifically it's shown some optimality properties of the fixed points of the cavity equations, including optimality under the two post-processing procedures defined in MGW (namely *Strong Pruning* and *Minimum Spanning Tree*) and global optimality of the MPA solution in some limit cases.



In section 2.2 there is an introduction to combinatorial optimization problems and case complexity. Then, in section 2.3 the prize collecting steiner tree is defined, and it is shown how to make local the global constraint of the problem. The derivation of the Max-Sum equations for the PCST is made in sec. 2.4. In sec. 2.5 concurrent algorithms are presented. Numerical results are shown in sec. 2.6, and then in the last sec. 2.7 some optimality properties of the solutions found by `msgsteiner` are proved.

## 2.2 Combinatorial Optimization Problems

The Combinatorial Optimization problems (COPs) are a large family of problems that belong to the general class of computational problems. In general a COP consists in finding an element of a finite set which minimizes an easy-to-evaluate cost function. The reasons why this class of problems is interesting are:

- The COPs is one of the most fundamental class of problems within computational-complexity theory.
- They are ubiquitous both in applications and in pure sciences.
- There are strong connection between statistical physics and COPs.

The definition of a COP is easy. We have a finite set  $\Xi$  of allowed configurations of a systems and a cost function  $E$  (or energy) defined on this set and taking real values. We want to know what is the optimal configuration  $C$  that minimize the cost function  $E(C)$ . The connection to statistical physics is straightforward. If we define a measure in this way:

$$P_{\beta}(C) = \frac{e^{-\beta E(C)}}{Z_{\beta}}, \quad Z_{\beta} = \sum_{C \in \Xi} e^{-\beta E(C)} \quad (2.1)$$

where  $\beta$  is a fictitious inverse temperature, and we take the limit  $\beta \rightarrow \infty$  we recover the ground state of the system, namely the configuration that minimize the energy or cost function  $E$ .

### 2.2.1 Complexity scenario and NP class

COPs can be classified by own complexity. For instance a COP is considered *feasible* if it belong to the  $P$  (Polynomial time) class. The  $P$  class contains the problems that require a time to be solved that grows at least polynomially with the problem size. The  $P$  class belong to the NP class. A problem belongs to NP class if it is possible to test, in polynomial time, whether a specific presumed solution is correct. In NP exists another class, the NP-complete one. The COPs that belong to the latter class have the property that all NP problems can be mapped, in polynomial time, to them. Typically the computer time needed to solve a NP-complete problem grows exponentially with the problem size, in the worst case. Then there is the NP-hard class that contains problems hard as the NP-complete ones<sup>1</sup> that don't belong to NP class. The COP described in this thesis, the prize collecting steiner tree, belong to the NP-hard class.

## 2.3 Prize collecting steiner tree (PCST)

The PCST problem can be stated in general terms as the problem of finding a connected subgraph of minimum cost. It has applications in many areas ranging from biology, e.g. finding protein associations in cell signaling [42–44], to network technologies, e.g. finding optimal ways to deploy fiber optic and heating networks for households and industries [45].

### 2.3.1 definition

In the following we will describe the Prize-Collecting Steiner Tree problem on Graphs (see e.g. [47, 49]).

Given a network  $G = (V, E)$  with positive (real) weights  $\{c_e : e \in E\}$  on edges and  $\{b_i : i \in V\}$  on vertices, consider the problem of finding the connected sub-graph  $G' = (V', E')$  that minimizes the cost or energy function  $H(V', E') = \sum_{e \in E'} c_e -$

---

<sup>1</sup>Each problems in NP-hard can be mapped in polynomial time to a problem in NP-complete.

$\lambda \sum_{i \in V'} b_i$ , i.e. to compute the minimum:

$$\min_{\substack{E' \subseteq E, V' \subseteq V \\ (V', E') \text{ connected}}} \sum_{e \in E'} c_e - \lambda \sum_{i \in V'} b_i. \quad (2.2)$$

It can be easily seen that a minimizing sub-graph must be a tree (links closing cycles can be removed, lowering  $H$ ). The parameter  $\lambda$  regulates the tradeoff between the edge costs and vertices prizes, and its value has the effect to determine the size of the subgraph  $G'$ : for  $\lambda = 0$  the empty subgraph is optimal, whereas for  $\lambda$  large enough the optimal subgraph includes all nodes.

This problem is known to be NP-hard. To solve it we will use a variation of a very efficient heuristics based on belief propagation developed on [43, 44, 50] that is known to be exact on some limit cases [50, 51]. We will partially extend the results in [51] to a more general PCST setting.

### 2.3.2 Rooted, depth bounded PCST and forests

We will deal with a variant of the PCST called  $D$ -bounded rooted PCST ( $D$ -PCST). This problem is defined by a graph  $G$ , an edge cost matrix  $c$  and prize vector  $b$  along with a selected “root” node  $r$ . The goal is to find the  $r$ -rooted tree with maximum depth  $D$  of minimum cost, where the cost is defined as in (2.2). A general PCST can be reduced to  $D$ -bounded rooted PCST by setting  $D = |V|$  and probing with all possible rootings, slowing the computation by a factor  $|V|$  (we will see later a more efficient way of doing it). A second variant which we will consider is the so-called  $R$  multi-rooted  $D$ -bounded Prize Collecting Steiner Forest ( $(R, D)$ -PCSF). It consists of is a natural generalization of the previous problem: a subset  $R$  of “root” vertices is selected, and the scope is to find a forest of trees of minimum cost, each one rooted in one of the preselected root nodes in  $R$ .

### 2.3.3 Local constraints

The cavity formalism can be adopted and made efficient if the global constraints, which may be present in the problem, can be written in terms of local constraints,

see cap ?. In the PCST case the global constraint is connectivity which can be made local as follows.

We start with the graph  $G = (V, E)$  and a selected *root* node  $r \in V$ . To each vertex  $i \in V$  there is an associated couple of variables  $(p_i, d_i)$  where  $p_i \in \partial i \cup \{*\}$ ,  $\partial i = \{j : (ij) \in E\}$  denotes the set of neighbors of  $i$  in  $G$  and  $d_i \in \{1, \dots, D\}$ . Variable  $p_i$  has the meaning of the parent of  $i$  in the tree (the special value  $p_i = *$  means that  $i \notin V'$ ), and  $d_i$  is the auxiliary variable describing its distance to the root node (i.e. the *depth* of  $i$ ). To correctly describe a tree, variables  $p_i$  and  $d_i$  should satisfy a number of constraints, ensuring that depth decreases along the tree in direction to the root (the root node must be treated separately), i.e.  $p_i = j \Rightarrow d_i = d_j + 1$ . Additionally, nodes that do not participate to the tree ( $p_i = *$ ) should not be parent of some other node, i.e.  $p_i = j \Rightarrow p_j \neq *$ . Note that even though  $d_i$  variables are redundant (in the sense that they can be easily computed from  $p_j$  ones), they are crucial to maintain the locality of the constraints. For every ordered couple  $i, j$  such that  $(ij) \in E$ , we define  $f_{ij}(p_i, d_i, p_j, d_j) = \mathbb{1}_{p_i=j \Rightarrow d_i=d_j+1 \wedge p_j \neq *} = 1 - \delta_{p_i,j} (1 - \delta_{d_i,d_j+1} (1 - \delta_{p_j,*}))$  (here  $\delta$  is the Kroenecker delta). The condition of the subgraph to be a tree can be ensured by imposing that  $g_{ij} = f_{ij}f_{ji}$  has to be equal to one for each edge  $(ij) \in E$ . If we extend the definition of  $c_{ij}$  by  $c_{i*} = \lambda b_i$ , then (except for an irrelevant constant additive term), the minimum in (2.2) equals to:

$$\min \{ \mathcal{H}(\mathbf{p}) : (\mathbf{d}, \mathbf{p}) \in \mathcal{T} \}, \quad (2.3)$$

where  $\mathbf{d} = \{d_i\}_{i \in V}$ ,  $\mathbf{p} = \{p_i\}_{i \in V}$ ,  $\mathcal{T} = \{(\mathbf{d}, \mathbf{p}) : g_{ij}(p_i, d_i, p_j, d_j) = 1 \forall (ij) \in E\}$  and

$$\mathcal{H}(\mathbf{p}) \equiv \sum_{i \in V} c_{ip_i}. \quad (2.4)$$

This new expression for the energy accounts for the sum of taken edge costs plus the sum of uncollected prizes and has the advantage of being non-negative.

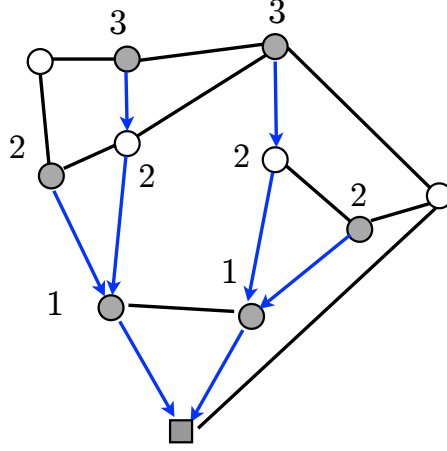


FIGURE 2.1: (Color online) A schematic representation of the Prize Collecting Steiner Tree problem and its local representation. Numbers next to the nodes are the distances (depths) from the root node (square node). The prize value is proportional to the darkness of the nodes. Arrows are the pointers from node to node. Distances and pointers are used to define the connectivity constraints which appear in the message-passing equations.

## 2.4 Derivation of the message-passing cavity equations

The algorithmic scheme we propose originates from the cavity method describe in the sec[1.2]. The starting point for the equations is the Boltzmann-Gibbs distribution:

$$P(\mathbf{d}, \mathbf{p}) = \frac{\exp(-\beta \mathcal{H}(\mathbf{p}))}{Z_\beta}, \quad (2.5)$$

where  $(\mathbf{d}, \mathbf{p}) \in \mathcal{T}$ ,  $\beta$  is a positive parameter (a fictitious inverse temperature), and  $Z_\beta$  is a normalization constant (the partition function). In the limit  $\beta \rightarrow \infty$  this probability concentrates on the configurations which minimize  $\mathcal{H}$ . Given  $i, j \in V$ , the message  $P_{ji}(d_j, p_j)$  is defined as the marginal distribution  $\sum_{(d_k, p_k)_{k \in V \setminus \{j, i\}}} P_{G^{(i)}}(\mathbf{d}, \mathbf{p})$  on a graph  $G^{(i)}$ , which equals to graph  $G$  minus node  $i$  and all its edges. The BP equations are derived:

$$P_{ji}(d_j, p_j) \propto e^{-\beta c_{jp_j}} \prod_{k \in \partial j \setminus i} Q_{kj}(d_j, p_j) \quad (2.6)$$

$$Q_{kj}(d_j, p_j) \propto \sum_{d_k} \sum_{p_k} P_{kj}(d_k, p_k) g_{jk}(d_k, p_k, d_j, p_j). \quad (2.7)$$

This assumption is correct if  $G$  is a tree, in which case (2.6)-(2.7) are exact and have a unique solution (see e.g. chapter 14.2 of [27]). The solutions of equations (2.6)-(2.7) are searched through iteration: substituting (2.7) onto (2.6) and giving a time index  $t+1$  and  $t$  to the cavity marginals in respectively the left and right hand side of the resulting equation, this system is iterated until numerical convergence is reached. On a fixed point, the BP approximation to the marginal is computed as

$$P_j(d_j, p_j) \propto e^{-\beta c_{jp_j}} \prod_{k \in \partial j} Q_{kj}(d_j, p_j). \quad (2.8)$$

### 2.4.1 Max-sum: $\beta \rightarrow \infty$ limit

In order to take the  $\beta \rightarrow \infty$  limit, (2.7) can be rewritten in terms of “cavity fields”

$$\psi_{ji}(d_j, p_j) = \beta^{-1} \log P_{ji}(d_j, p_j) \quad (2.9)$$

$$\phi_{kj}(d_j, p_j) = \beta^{-1} \log Q_{kj}(d_j, p_j). \quad (2.10)$$

The BP equations take the so-called Max-sum form:

$$\psi_{ji}(d_j, p_j) = -c_{jp_j} + \sum_{k \in \partial j \setminus i} \phi_{kj}(d_j, p_j) + C_{ji} \quad (2.11)$$

$$\phi_{kj}(d_j, p_j) = \max_{p_k, d_k: g_{jk}(d_k, p_k, d_j, p_j)=1} \psi_{kj}(d_k, p_k), \quad (2.12)$$

where  $C_{ji}$  is an additive constant chosen to ensure  $\max_{d_j, p_j} \psi_{ji}(d_j, p_j) = 0$

Computing the right side of (2.12) is in general too costly in computational terms. Fortunately, the computation can be carried out efficiently by breaking up the set over which the max is computed into smaller (possibly overlapping) subsets. We define

$$A_{kj}^d = \max_{p_k \neq j, *} \psi_{kj}(d, p_k) \quad (2.13)$$

$$B_{kj}^d = \psi_{kj}(d, *) \quad (2.14)$$

$$C_{kj}^d = \psi_{kj}(d, j). \quad (2.15)$$

Equation (2.12) can now be rewritten as:

$$A_{ji}^d = \sum_{k \in \partial j \setminus i} E_{kj}^d + \max_{k \in \partial i \setminus j} \{-c_{jk} - E_{kj}^d + A_{kj}^{d-1}\} \quad (2.16)$$

$$B_{ji} = -c_{j*} + \sum_{k \in \partial j \setminus i} D_{kj} \quad (2.17)$$

$$C_{ji}^d = -c_{ji} + \sum_{k \in \partial j \setminus i} E_{kj}^d \quad (2.18)$$

$$D_{ji} = \max \left( \max_d A_{ji}^d, B_{ji} \right) \quad (2.19)$$

$$E_{ji}^d = \max \left( C_{ji}^{d+1}, D_{ji} \right). \quad (2.20)$$

Using some simple efficiency tricks including computing  $\sum_{k \in \partial j \setminus i} E_{kj}^d$  as  $\sum_{k \in \partial j} E_{kj}^d - E_{ki}^d$ , the computation of the right side of (2.16)-(2.20) for all  $i \in \partial j$  can be done in a time proportional to  $D|\partial j|$ , where  $D$  is the depth bound. The overall computation time is then  $O(|E|D)$  per iteration.

### 2.4.2 Total fields

In order to identify the minimum cost configurations, we need to compute the total marginals, i.e. the marginals in the case in which no node has been removed from the graph. Given cavity fields, the total fields  $\psi_j(d_j, p_j) = \lim_{\beta \rightarrow \infty} \beta^{-1} \log P_j(d_j, p_j)$  can be written as:

$$\psi_j(d_j, p_j) = -c_{jp_j} + \sum_{k \in \partial j} \phi_{kj}(d_j, p_j) + C_j, \quad (2.21)$$

where  $C_j$  is again an additive constant that ensures  $\max_{d_j, p_j} \psi_j(d_j, p_j) = 0$ . In terms of the above quantities we find  $\psi_j(d_j, i) = F_{ji}^d \stackrel{\text{def}}{=} \sum_{k \in \partial j} E_{kj}^d + (-c_{ij} - E_{ji}^d + A_{ji}^{d-1})$  if  $i \in \partial j$  and  $\psi_j(d_j, *) = G_j \stackrel{\text{def}}{=} -c_{j*} + \sum_{k \in \partial j} D_{kj}$ . The total fields can be interpreted as (the Max-Sum approximation to) the relative negative energy loss of choosing a given configuration for variables  $p_j, d_j$  instead of their optimal choice, i.e.  $\psi_j(d_j, p_j) = \min \{\mathcal{H}(\mathbf{p}') : (\mathbf{d}', \mathbf{p}') \in \mathcal{T}\} - \min \{\mathcal{H}(\mathbf{p}') : (\mathbf{d}', \mathbf{p}') \in \mathcal{T}, d_j = d'_j, p_j = p'_j\}$ . In particular, in absence of degeneracy, the maximum of the field is attained for values of  $p_j, d_j$  corresponding to the optimal energy. In our simulations, the energies computed always correspond to the tree obtained by maximizing the total fields in this way.

### 2.4.3 Iterative dynamics and reinforcement

Equations (2.16)-(2.20) can be thought as a fixed-point equation in a high dimensional euclidean space. This equation could be solved by repeated iteration of the quantities  $A, B$ , and  $C$  starting from an arbitrary initial condition, simply by adding an index  $(t+1)$  to  $A, B, C$  in the left-hand side of (2.12) and index  $(t)$  to all other instances of  $A, B, C, D, E$ .

This system converges in many cases. When it does not converge, the *reinforcement* is of help [52]. The resulting new Max-sum equations become:

$$A_{ji}^d(t+1) = \sum_{k \in \partial j \setminus i} E_{kj}^d(t) + \quad (2.22)$$

$$+ \max_{k \in \partial j \setminus i} \{-c_{jk} - E_{kj}^d(t) + A_{kj}^{d-1}(t) + \gamma_t F_{jk}^d(t)\} \quad (2.23)$$

$$B_{ji}(t+1) = -c_{j*} + \sum_{k \in \partial j \setminus i} D_{kj}(t) + \gamma_t G_j(t) \quad (2.24)$$

$$C_{ji}^d(t+1) = -c_{ji} + \sum_{k \in \partial j \setminus i} E_{kj}^d(t) + \gamma_t F_{ji}^d(t) \quad (2.25)$$

$$D_{ji}(t) = \max \left\{ \max_d A_{ji}^d(t), B_{ji}(t) \right\} \quad (2.26)$$

$$E_{ji}^d(t) = \max \left\{ C_{ji}^{d+1}(t), D_{ji}(t) \right\} \quad (2.27)$$

$$G_j(t+1) = -c_{j*} + \sum_{k \in \partial j} D_{kj}(t) + \gamma_t G_j(t) \quad (2.28)$$

$$F_{ji}^d(t+1) = \sum_{k \in \partial j} E_{kj}^d(t) + (-c_{ji} - E_{ij}^d(t) + A_{ij}^{d-1}(t)) + \quad (2.29)$$

$$+ \gamma_t F_{ji}^d(t). \quad (2.30)$$

In our experiments, the equations converge for a sufficiently large  $\gamma_t$ . The strategy we adopted is, when the equations do not converge, to start with  $\gamma_t = 0$  and slowly increase it until convergence in a linear regime  $\gamma_t = t\rho$  (although other regimes are possible). The number of iterations is then found to be inversely dependent on the parameter  $\rho$ . This strategy could be interpreted as using time-averages of the MS marginals when the equations do not converge to gradually bootstrap the system into an (easier to solve) system with sufficiently large external fields. A C++ implementation of these equations can be found (in source form) on [53].



Note that the cost matrix  $(c_{ij})$  need not to be symmetric, and the same scheme could be used for directed graphs (using  $c_{ji} = \infty$  if  $(i, j) \in E$  but  $(j, i) \notin E$ ).

#### 2.4.4 Root choice

The PCST formulation given in the introduction is unrooted. The MS equations on the other hand, need a predefined root. One way of reducing the unrooted problem to a rooted problem is to solve  $N = |V|$  different problems with all possible different rooting, and choose the one of minimum cost. This unfortunately adds a factor  $N$  to the time complexity. Note that in the particular case in which some vertex has a large enough prize to be necessarily included in an optimal solution (e.g.  $\lambda b_i > \sum_{e \in E} c_e$ ), this node can simply be chosen as root.

We have devised a more efficient method for choosing the root in the general case, which we will now describe. Add an extra new node  $r$  to the graph, connected to every other node with identical edge cost  $\mu$ . If  $\mu$  is sufficiently large, the best energy solution is the (trivial) tree consisting in just the node  $r$ . Fortunately, a solution of the MS equations on this graph gives additional information: for each node  $j$  in the original graph, the marginal field  $\psi_j$  gives the relative energy shift of selecting a given parent (and then adjusting all other variables in the best possible configuration). Now for each  $j$ , consider the positive real value  $\alpha_j = -\psi_j(1, r)$ , that corresponds with the best attainable energy, constrained to the condition that  $r$  is the parent of  $j$ . If  $\mu$  is large enough, this energy is the energy of a tree in which only  $j$  (and no other node) is connected to  $r$  (as each of these connections costs  $\mu$ ). But these trees are in one to one correspondence with trees rooted at  $j$  in the original graph. The smallest  $\alpha_j$  will thus identify an optimal rooting.

Unfortunately the information carried by these fields is not sufficient to build the optimal tree. Therefore one needs to select the best root  $j$  and run the MS equations a second time on the original graph using this choice.

## 2.5 other method

We compared the performance of `msgsteiner` with three different algorithms: two that employ an integer linear programming strategy to find an optimal subtree,

namely the Lagrangian Non Delayed Relax and Cut (LNDRC) [54] and branch-and-cut (DHEA) [48], and a modified version of the Goemans and Williamson algorithm (MGW)[47].

### 2.5.1 integer linear programming

The goal of Integer Linear Programming (ILP) is to find an integer vector solution  $x^* \in \mathbb{Z}^n$  such that:

$$c^T x^* = \min\{c^T x^* \mid Ax \geq b, x \in \mathbb{Z}^n\}, \quad (2.31)$$

where a matrix  $A \in \mathbb{R}^{m \times n}$  and vector  $b \in \mathbb{R}^m$  and  $c \in \mathbb{R}^n$  are given. Many graph problems can be formulated as an integer linear programming problem [55]. In general, solving (2.31) with  $x^* \in \mathbb{Z}$  is NP-Complete. The standard approach consists in solving (2.31) for  $x^* \in \mathbb{R}$  (a relaxation of the original problem) and use the solution as a guide for some heuristics or complete algorithm for the integer case. The relaxed problem can be solved by many classical algorithms, like the Simplex Method or Interior Point methods [56]. In order to map the PCST problem in a ILP problem we introduce a variable vector  $z \in \{0, 1\}^E$  and  $y \in \{0, 1\}^V$  where the component for an edge in  $E$  or for a vertex in  $V$  is one if it is included in the solution and zero otherwise. Now (2.2) can be written as

$$H = \sum_{e \in E} c_e z_e - \sum_{i \in V} b_i y_i, \quad (2.32)$$

and the constraints  $Ax \geq b$  in (2.31), that are used to enforce that induced subgraph is a tree, generally involve all or most of the variables  $z$  and  $y$ . Roughly speaking  $Ax \geq b$  in (2.31) defines a bounded volume in the space of parameters, i.e. a polytope. The equation to minimize is linear (2.31) so the minimum is on a vertex of this polytope. In general for hard problems, in order to ensure that each vertex of the polytope (or more in particular just the optimal vertex) is integer, a number of extra constraints that grows exponentially with the problem size may be needed [55].

DHEA and LNRDC use different techniques to tackle the problems of the enormous number of resulting constraints. Both programs are able in principle to prove the optimality of the solution (if given sufficient/exponential time), and when it

is not the case they are able to give a lower bound for the value of the optimum given by the the optimum of the relaxed problem.

### 2.5.2 Goemans-Williamson

The MGW algorithm is based on the primal-dual method for approximation algorithms [46]. The starting point is still the ILP formulation of the problem (2.31), but it employs a controlled approximation scheme that enforces the cost of any solution to be at most twice as large as the optimum one. In addition, MGW implements two different post-processing strategies, namely a pruning scheme that is able to eliminate some nodes while lowering the cost, and the computation of the minimum spanning tree in order to find an optimal rewiring of the same set of nodes. The overall running time is  $\mathcal{O}(n^2 \log n)$ . A complete description is available in [46].

## 2.6 Computational Experiments

### 2.6.1 Instances

Experiments were performed on several classes of instances:

- C, D and E available at [57] and derived from the Steiner problem instances of the OR-Library [58]. This set of 120 instances was previously used as benchmark for algorithms for the PCST[58]. The solutions of these instances were obtained with the algorithms[48, 54]. The class C, D, E have respectively 500, 1000, 2000 nodes and are generated at random, with average vertex degree: 2.5, 4, 10 or 50. Every edge cost is a random integer in the interval  $[1, 10]$ . There are either 5, 10,  $n/6$ ,  $n/4$  or  $n/2$  vertices with prizes different from zero and random integers in the interval  $[1, \text{maxprize}]$  where *maxprize* is either 10 or 100. Thus, each of the classes C, D, E consists of 40 graphs.
- K and P available at [57]. These instances are provided in [47]. In the first group instances are unstructured. The second group includes random geometric instances designed to have a structure somewhat similar to street

maps. Also the solution of these instances were found with the algorithms [48, 54].

- H are the so-called hypercubes instances proposed in [59]. These instances are artificially generated and they are very difficult instances for the Steiner tree problem. Graphs are  $d$ -dimensional hypercubes with  $d \in 6, \dots, 12$ . For each value of  $d$ , the corresponding graph has  $2^d$  vertices and  $d \cdot 2^{d-1}$  edges. We used the prized version of these instances defined in [48]. For almost all instances in this class the optimum is unknown.
- i640 are the so-called incidence instances proposed in [60] for the Minimum Steiner Tree problem. These instances have 640 nodes and only the nodes in a subset  $K \subseteq V$  have prizes different from zero (in the original problem these were terminals). The weight on each edge  $(i, j)$  is defined with a sample  $r$  from a normal distribution, rounded to an integer value with a minimum outcome of 1 and maximum outcome of 500, i.e.,  $c_{ij} = \min\{\max\{1, \text{round}(r)\}, 500\}$ . However, to obtain a graph that is much harder to reduce by preprocessing techniques three distributions with a different mean value are used. Any edge  $(i, j)$  is incident to none, to one, or to two vertices in subset  $K$ . The mean of  $r$  is 100 for edges  $(i, j)$  with  $i, j \notin K$ , 200 on edges with one end vertex in  $K$ , and 300 on edges with both ends in  $K$ . Standard deviation for each of the three normal distributions is 5. In order to have prizes also on vertices we extracted uniformly from all integer in the interval between 0 and  $4 \cdot \max_{edge}$  where  $\max_{edge}$  is the maximum value of edges in the samples considered. There are 20 variants combining four different number of vertices in  $K$  (rounding to the integer value  $\lfloor \cdot \rfloor$ ):  $|k| = \lfloor \log_2 |V| \rfloor$ ,  $\lfloor \sqrt{|V|} \rfloor$ ,  $\lfloor 2\sqrt{|V|} \rfloor$ , and  $\lfloor |V|/4 \rfloor$  with five edge number:  $|E| = \lfloor 3|V|/2 \rfloor$ ,  $2|V|$ ,  $\lfloor |V| \log |V| \rfloor$ ,  $\lfloor 2|V| \log |V| \rfloor$ , and  $\lfloor |V|(|V| - 1)/4 \rfloor$ . Each variant is drawn five times, giving 100 instances.
- Class R. The last class of samples are  $G(n, p)$  random graphs with  $n$  vertices and independent edge probability  $p = (2\nu)/(n - 1)$ . The parameter  $\nu$  is the average node degree, that was chosen as  $\nu = 8$ . The weight on each edge  $(i, j)$  can take three different value, 1, 2 and 4, with equal probability  $\frac{1}{3}$ . Node prizes were extracted uniformly in the interval  $[0, 1]$ . We generated different graphs with four different values of  $\lambda$  ( $\lambda = 1.2, 1.5, 2$  or  $3$ ), see (2.2), in order to explore different regimes of solution sizes. We find that the average number of nodes that belong to the solution for  $\lambda = 1.2, 1.5, 2$

and 3 are respectively about 14%, 33%, 51%, 67% of the total nodes in the graph. We have created twelve instances of different size for the four class of random graph, from  $n = 200$  up to  $n = 4000$  nodes. For each parameter set we generated ten different realizations. The total number of samples is 480.

The `msgsteiner` algorithm was implemented in C++ and run on a single core of an AMD Opteron Processor 6172, 2.1GHz, 8 Gb of RAM, with Linux, g++ compiler, -O3 flag activated. A C++ implementation of these equations can be found in source form on [53]. The executable of DHEA is available in [57], and in order to compare the running time we ran DHEA and `msgsteiner` on the same workstation. The executable of LNDRC and MGW programs was not available. We implemented the non-rooted version of MGW to compare only the optimum on the random graph instances.

## 2.6.2 Results

We analyzed two numeric quantities: the time to find the solution, and the gap between the cost of the solution and the best known lower bound (or the optimum solution when available) typically found with programs based on linear programming. The gap is defined as  $gap = 100 \cdot \frac{Cost - LowerBound}{lowerBound}$ .

In Table (2.1) we show the comparison between `msgsteiner` and the DHEA program. DHEA is able to solve exactly K, P and C, D, E instances. The worst performance of `msgsteiner` is on the K class, where the average gap is about 2.5%. In this class the average solution is very small as it comprises only about 4.4% of total nodes of the graph. `msgsteiner` seems to have most difficulty with small subgraphs. `msgsteiner` is able to find solutions very close to the optimum for the P class, that should be model a street network. `msgsteiner` is also able to find solutions very close to the optimum, with a gap inferior to 0.025% on the C, D, and E classes.

In Figure (2.2) we show the gap of `msgsteiner` and MGW from the optimum values found by the DHEA program in the class R. `msgsteiner` gaps are almost negligible (always under 0.05%) and tend to zero when the size grows. MGW gaps instead are always over 1%. For intermediate size of solutions trees the gaps of MGW are over 3%.

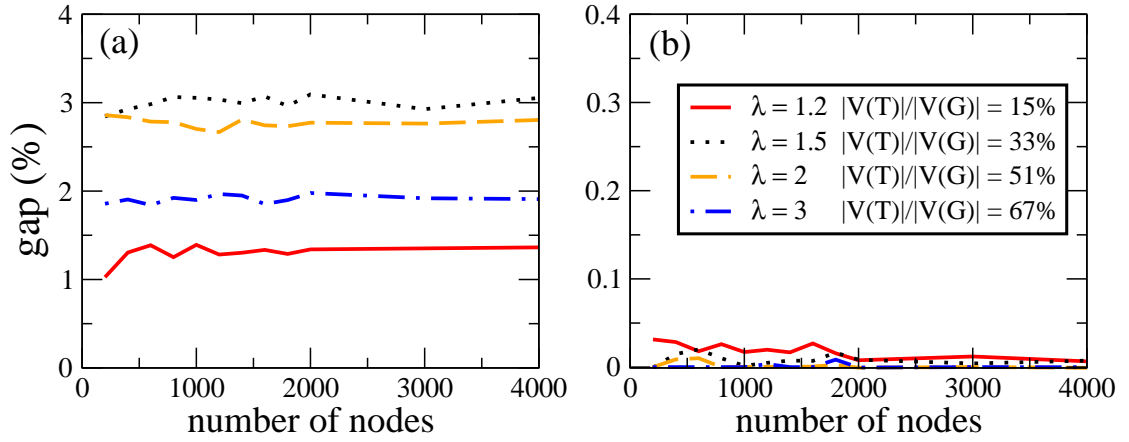


FIGURE 2.2: (Color online) Gap from the optimum found by DHEA program of respectively (a) MGW and (b) `msgsteiner` as a function of the number of nodes. `msgsteiner` gaps are always under 0.05%. MGW gaps are always over 1% and for intermediate sizes of the solution tree the gaps of MGW are over 3%.

In Figure (2.3) we show the running time for the class R, with increasing solution tree size. In general we observe that the running time of `msgsteiner` grows much slower than the one of DHEA for increasing number of nodes in the graph and `msgsteiner` largely outperforms DHEA in computation time for large instances; furthermore the differences between the algorithms became specially and large for large expected tree solution. In at least one case DHEA could not find the optimum solution within the required maximum time and the `msgsteiner` solution was slightly better.

The class i640 consists in graphs with varying number of edges and nodes, and a varying number of nodes with non-zero prize. We define  $K$  as the subset of nodes with non-zero prize. Table (2.2) shows, for each type of graph, the average time and the average gap on five different realizations of the graphs for `msgsteiner` and DHEA algorithms. We set the time limit to find a solution of DHEA to 2000 seconds. We observe that DHEA obtains good performance in terms of the optimality of the solution when the size of subset  $K$  is small. `msgsteiner` finds better result than DHEA when the size of  $K$  is sufficiently large, within a time of one or two order of magnitude smaller. Moreover DHEA seems to have difficulty to find reasonable good solution when the graph have high connectivity.

We show in Table (2.3) a comparison between `msgsteiner`, LNDRC [54] and DHEA. The results and running time of LNDRC are taken from [54]. The computer reportedly used for the optimization is comparable with ours. We have

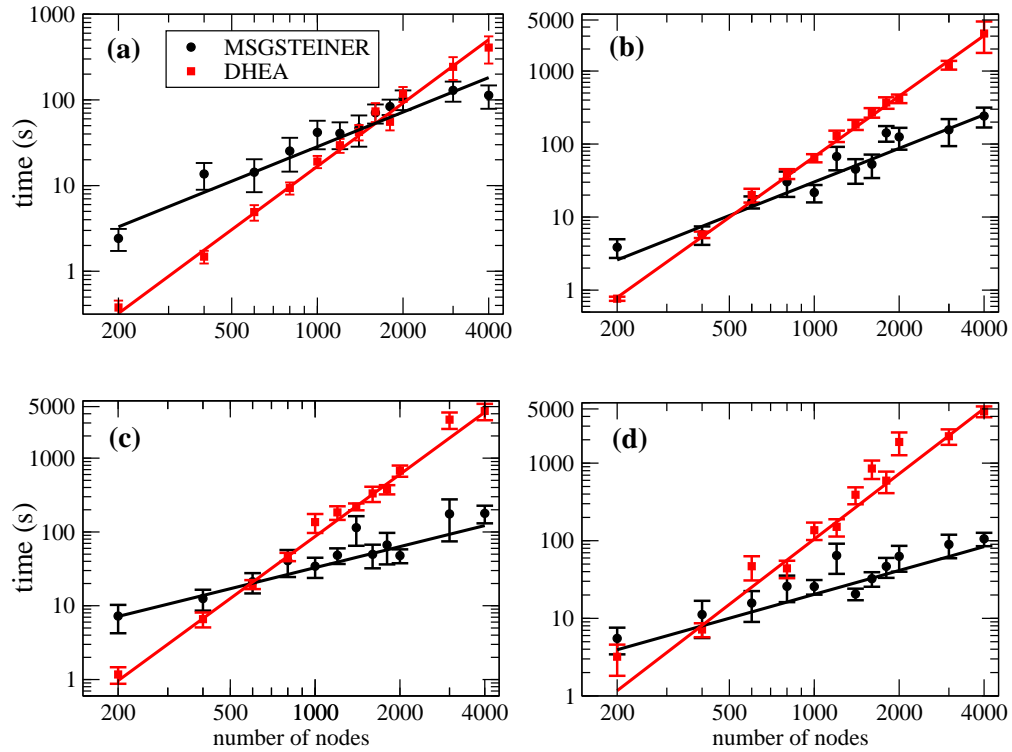


FIGURE 2.3: (Color online) Result on the random graphs class R. Points correspond to the running time of `msgsteiner` and DHEA versus graph size. The four cases show how running time behavior depends on size of the expected solution tree. The plots (a), (b), (c), (d) have respectively  $\lambda = 1.2, 1.5, 2, 3$  and average value of the fraction of nodes that belongs to the solution respectively  $|V(T)|/|V(G)| = 0.14, 0.33, 0.51, 0.67$ . The quantities shown in figure are averaged over ten different realizations. Data are fitted with function  $y = ax^b$ . The  $b$  values found for DHEA are for (a), (b), (c), (d) respectively : 2.4, 2.8, 2.8, 2.8. BP performance is as expected roughly linear in the number of vertices. The fitted  $b$  parameters are for (a), (b), (c), (d) respectively: 1.5, 1.3, 1, 1. For instances that are large enough, the running time of `msgsteiner` is smaller than the one of DHEA and the difference increases with the expected solution tree.

imposed to DHEA a time limit of 6000 seconds and we show two results of `msgsteiner` with different values of the reinforcement parameter. The lower bound is taken from [54]. In almost all instance `msgsteiner` obtains better results, both in time and in quality of solution. The difference is accentuated for large instances. As expected, decreasing the reinforcement parameter allows to find lower costs at the expense of larger computation times.

Group	MS gap	MS time (s)	DHEA gap	DHEA time (s)	Size Sol
K	2.62%	6.51	0.0%	127.97	4.4%
P	0.46%	2.31	0.0%	0.18	31.4%
C	0.006%	16.24	0.0%	2.30	20.2%
D	0.005%	35.06	0.0%	16.12	20.2%
E	0.024%	305.49	0.0%	1296.11	26.4%

TABLE 2.1: Results class KPCDE

Name	time MS	time DHEA	gap MS (%)	gap DHEA (%)
0-0	0.8	0.2	1.3	0
0-1	2.5	4.2	1.0	0
0-2	100.8	226.6	1.4	0
0-3	1.2	0.3	0.05	0
0-4	37.3	72.8	1.8	0
1-0	1.0	0.85	0.3	0
1-1	2.6	1060.1	1.2	1.5
1-2	90.6	1133.8	0.7	0.2
1-3	1.5	3.8	0.8	0
1-4	33.7	2000.0	1.8	7.8
2-0	0.8	0.7	0.1	0
2-1	4.3	2000.0	2.2	11.6
2-2	149.7	2011.1	0.8	14.8
2-3	1.2	12.0	0.2	0
2-4	39.2	2001.0	1.9	11.2
3-0	1.1	2.4	0.3	0
3-1	3.9	2001.0	1.7	5.6
3-2	112.6	2015.1	0.8	4.9
3-3	1.6	145.3	0.2	0
3-4	33.1	2000.5	1.2	59.9
mean	31.0	834.6	1.0	5.9

TABLE 2.2: Results i640 class

### 2.6.3 discussion

In this chapter we compared `msgsteiner`, with two state-of-the art algorithms for the Prize-Collecting Steiner Problem. The Cavity Theory is expected to give asymptotically exact results on many ensembles of random graphs, so we expected it to give better performance for large instances. The comparison was performed both on randomly-generated graphs and existing benchmarks. We observed that `msgsteiner` finds better costs in significantly smaller times for many of the instances analyzed, and that this difference in time and quality grew with the size of the instances and their solution. We find these results encouraging in views of



Name	MS(-5)		MS(-3)		LNDRC		DHEA	
	gap(%)	time(s)	gap(%)	time(s)	gap(%)	time(s)	gap(%)	time(s)
6p	2.2	3.5	2.6	0.6	4.2	0.5	2.2	21.3
6u	1.5	6.4	4.3	0.7	4.3	0.5	1.5	0.4
7p	2.3	90.2	3.9	1.7	7.7	1.5	2.3	6000.3
7u	2.2	134.1	2.2	1.8	3.6	1.2	2.2	596.4
8p	2.4	255.5	3.4	3.8	7.1	5.2	2.3	6004.2
8u	1.8	351.1	3.3	4.9	7.5	4.1	3.3	6000.9
9p	1.8	555.6	2.3	10.8	8.6	16.1	22.1	6000.0
9u	1.9	775.8	3.3	11.1	6.2	13.1	Not Found	6000.4
10p	1.7	1761.9	1.7	28.0	10.4	114.4	31.3	6000.5
10u	2.7	2468.4	2.7	32.2	7.7	59.8	Not Found	6000.6
11p	1.5	972.3	1.6	49.3	11.6	630.0	Not Found	6003.1
11u	2.2	5632.8	2.6	71.9	9.0	360.6	Not Found	6001.5
12p	1.5	4970.8	1.6	121.4	11.3	3507.7	Not Found	6009.8
12u	2.0	4766.7	2.4	174.1	10.0	1915.7	Not Found	6002.3
mean	2.0	1624.7	2.7	36.6	7.8	473.6	—	4760.1

TABLE 2.3: Results H class

future applications to problems in biology in which optimization of networks with millions of nodes may be necessary, in particular given the conceptual simplicity of the scheme behind `msgsteiner` (a simple fixed-point iteration).

## 2.7 Post-processing and optimality

For this section we will assume unbounded depth  $D$ . Results are not easily generalizable to the bounded- $D$  case. Results in this section apply to the non-reinforced MS equations ( $\gamma_t = 0$ ). The results here are based in construction of certain trees associated with the original graph and in the fact that MS/BP equations are always exact and have a unique solution on trees [51].

**Definition 1.** Let  $\{\psi_{ij}\}$  be a MS fixed-point (2.11)-(2.12), and let  $\mathbf{d}, \mathbf{p}$  be the decisional variables associated with this fixed point, i.e.  $(d_i^*, p_i^*) = \arg \max \psi_i(d_i, p_i)$  for the physical field  $\psi_i$  from (2.21). We will assume this maximum to be non degenerate. We will employ the *induced subgraph*  $S^*=(V^*, E^*)$  defined by  $V^* = \{i \in V : p_i^* \neq *\} \cup \{r\}$  and  $E^* = \{(i, p_i^*) : i \in V, p_i^* \in V\}$ . The cost of this subgraph is  $\mathcal{H}(S^*) = \mathcal{H}(\mathbf{p}) = \sum_{i \in V} c_{ip_i^*}$ .

The following local optimality property of the Max-Sum-induced solution will be proven in the appendix A. It states that a MS solution is no worse than any subgraph containing a subset of the nodes.

**Theorem 2.** *Given a MS fixed point  $\{\psi_{ij}\}$  on  $G$  (unbounded  $D$ ) with induced subgraph  $S^* = (V^*, E^*)$  and any subtree  $S' = (V', E') \subseteq G$  with  $V' \subseteq V^*$ , then  $\mathcal{H}(S^*) \leq \mathcal{H}(S')$*

This result has an easy generalization to loopy subgraphs:

**Corollary 3.** *With  $S^*$  as in Theorem 2, given any connected subgraph  $S' = (V', E') \subseteq G$  with  $V' \subseteq V^*$ , then  $\mathcal{H}(S^*) \leq \mathcal{H}(S')$ .*

*Proof.* Apply Theorem (2) to a spanning tree of  $S'$ . □

This trivially implies the following result of global optimality of the MS solution in a particular case:

**Corollary 4.** *With  $S^* = (V^*, E^*)$  as in Theorem 2, if  $V^* = V$  then  $\mathcal{H}(S^*) = PCST(G)$*

In [47], the MGW algorithm includes two additional methods to obtain a better PCST solution: StrongPrune and Minimum Spanning Tree (MST) maintaining the same vertex set. Both methods give a substantial improvement boost to the MGW candidate computed in the first phase. A natural question may arise, does any of these two methods may help to improve the solution of MS? The answer is negative in both cases, and it is a trivial consequence of Theorem (2).

**Corollary 5.**  $MST(V^*, E \cap (V^* \times V^*)) = \mathcal{H}(S^*)$

*Proof.* The minimum spanning tree of  $(V^*, E \cap (V^* \times V^*))$  satisfies the hypothesis of Theorem (2), so  $\mathcal{H}(S^*) \leq MST(V^*, E \cap (V^* \times V^*))$ . The converse inequality is trivially true due to the optimality of the MST. □

**Corollary 6.**  $\mathcal{H}(StrongPrune(S^*)) = \mathcal{H}(S^*)$

*Proof.* This is a consequence of the fact that  $V(StrongPrune(S^*)) \subseteq V(S^*) = V^*$  and thus Theorem (2) applies, implying  $\mathcal{H}(S^*) \leq \mathcal{H}(StrongPrune(S^*))$ . The opposite inequality  $\mathcal{H}(StrongPrune(F)) \leq \mathcal{H}(F)$  was proved in [47]. □

# Chapter 3

## Variational Quantum Cavity Method

### 3.1 Introduction

In this chapter it is presented a technique that allows to compute approximate low excite states of many body quantum systems. In general it is a computationally hard task even founding the ground-state in one-dimensional quantum systems. Exact diagonalization algorithms are very helpful but they are limited to small systems due to the exponential growth of the Hilbert space with the system size [61]. For larger systems one has to resort to approximation methods, e.g., the variational quantum Monte Carlo algorithms, to study the low-energy states of the Hamiltonian [62–64]. On the other hand, one can always obtain useful insights by studying some exactly solvable mean-field models [65].

The estimation of local expectation values is of central importance in classical and quantum statistical physics in order to infer the physical state of an interacting system. This is in general a computationally hard problem especially for disordered systems displaying glassy behaviors, where approximation algorithms based on the Monte Carlo sampling could be very time consuming. At least for mean-field systems and finite-connectivity models with a locally tree-like interaction graph, the cavity method, described in the first chapter, provides efficient message-passing algorithms.

In the work of [Ramezanpour](#) it was presented an approach that merges the technique explained in [1.4.1](#) for stochastic optimization problems with the variational principle of quantum mechanics, i.e. the fact that any trial wave function provides an upper bound for the ground-state energy. The problem of finding the optimal state can be recast as an optimization problem with an objective function that evaluates the energy of the system. The idea is to map the trial wave function on a Gibbs state of a classical system and to consider the set of couplings as variational parameters. Crucial in this respect is the choice of the classical systems that should be able to capture the quantum nature of the original system. Given the trial wave function, the average of the quantum Hamiltonian can be estimated by cavity method. When we work on quantum systems that have a structure different from trees, Hamiltonian averages computed by the cavity method is an estimation of the true one, so in this case the upper bound property of the Hamiltonian is lost. In the chapter [4](#) it is presented a different technique that solves the same optimization problems in a faster and more precise way.

In order to study to the lower excited state of a system we can use the same procedure described above imposing additionally the orthogonality conditions between lower-energy eigenstates of the Hamiltonian. The main point is to write the overlap between two wave functions in the Bethe approximation, which is asymptotically exact as long as the trial wave functions can be represented by classical systems of locally tree-like interaction graphs. This allows us to replace the global orthogonality condition with some local constraints on the variational parameters and the cavity messages in the Bethe expression for the overlap.

In section [3.2](#) the variational quantum cavity method [\[15\]](#) is presented. Then, in section [3.3](#) it is shown the method to find low energy excited states within the Bethe measure [\[17\]](#).

## 3.2 Ground state approximate wave function

Given a Hamiltonian  $H$  and a trial wave function  $|\psi(\mathbf{P})\rangle$ , we have  $\langle\psi(\mathbf{P})|H|\psi(\mathbf{P})\rangle \geq E_g$  where  $E_g$  is the ground-state energy of  $H$  and  $\mathbf{P}$  denotes a set of parameters characterizing the trial wave function. To find the optimal parameters we define

the following optimization problem:

$$\mathcal{Z} = \sum_{\mathbf{P}} e^{-\beta_{opt} \langle \psi(\mathbf{P}) | H | \psi(\mathbf{P}) \rangle}, \quad (3.1)$$

where eventually one is interested in the limit  $\beta_{opt} \rightarrow \infty$ . Note that  $\beta_{opt}$  is just the inverse of a fictitious temperature and has nothing to do with the physical temperature. In the equation (3.2) we have to perform a sum over a set of parameters that change the average value of a energy function defined, at least for systems addressed in this thesis, over discrete variables. This problems is very close to a stochastic optimization problems showed in sec 1.4.1, as well as the technics I will show in the next sections to solve it.

Assume  $H = H_0 + H_1$ , where  $H_0$  is diagonal in the orthonormal basis  $|\sigma\rangle$ ;  $H_0$  and  $H_1$  are Hermitian operators with real eigenvalues and orthonormal eigenvectors. The trial wave function is represented in this representation as  $|\psi(\mathbf{P})\rangle = \sum_{\sigma} \psi(\sigma; \mathbf{P}) |\sigma\rangle$ . The coefficients  $\psi(\sigma; \mathbf{P})$  are complex numbers, and  $|\psi(\sigma; \mathbf{P})|^2 = \rho(\sigma; \mathbf{P})$  is a normalized probability distribution over  $\sigma$ . The average energy can be written as

$$\langle \psi(\mathbf{P}) | H | \psi(\mathbf{P}) \rangle = \sum_{\sigma} |\psi(\sigma; \mathbf{P})|^2 [E_0(\sigma) + E_1(\sigma)] \quad (3.2)$$

$$= \sum_{\sigma} \rho(\sigma; \mathbf{P}) [E_0(\sigma) + E_1(\sigma)], \quad (3.3)$$

where

$$E_0(\sigma) \equiv \langle \sigma | H_0 | \sigma \rangle, \quad E_1(\sigma) \equiv \text{Re} \left\{ \sum_{\sigma'} \frac{\psi^*(\sigma'; \mathbf{P})}{\psi^*(\sigma; \mathbf{P})} \langle \sigma' | H_1 | \sigma \rangle \right\}. \quad (3.4)$$

We consider  $\rho(\sigma; \mathbf{P}) \equiv |\psi(\sigma; \mathbf{P})|^2$  as a probability measure in a classical system and compute the above average quantities within the Bethe approximation. For  $\psi(\sigma; \mathbf{P}) \propto \prod_{a \in \mathcal{E}_c} \phi_a(\sigma^{\partial a}; P_a)$ , the classical measure is given by  $\rho(\sigma; \mathbf{P}) \propto \prod_a |\phi_a(\sigma^{\partial a}; P_a)|^2$  with the set of classical interactions  $\mathcal{E}_c \equiv \{\phi_a(\sigma^{\partial a}) | a = 1, \dots, A\}$ . Here  $\partial a$  is the subset of variables that appear in  $\phi_a$ . We write the BP equation

(1.44) over this classical systems:

$$\mu_{i \rightarrow a}(\sigma_i) \propto \prod_{b \in \partial i \setminus a} \mu_{b \rightarrow i}(\sigma_i), \quad (3.5)$$

$$\mu_{a \rightarrow i}(\sigma_i) \propto \sum_{\boldsymbol{\sigma}^{\partial a \setminus i}} |\phi_a(\boldsymbol{\sigma}^{\partial a}; P_a)|^2 \prod_{j \in \partial a \setminus i} \mu_{j \rightarrow a}(\sigma_j). \quad (3.6)$$

The average of any local quantity like  $E_0$  and  $E_1$  can be written as a function of the BP cavity marginals (or messages). Therefore, the optimization problem reads

$$\mathcal{Z} = \sum_{\mathbf{P}} \sum_{\boldsymbol{\mu}} e^{-\beta_{opt} \langle E_0 \rangle_{\rho} - \beta_{opt} \langle E_1 \rangle_{\boldsymbol{\mu}}} \mathbb{1}_{BP}, \quad (3.7)$$

where the indicator function  $\mathbb{1}_{BP}$  ensures that the messages  $\boldsymbol{\mu}$  satisfy the BP equations, like shown for stochastic optimization problem in sec.1.4.1. In the case of multiple BP fixed points the above partition function would be concentrated on one minimum of the average energy for  $\beta_{opt} \rightarrow \infty$ . More accurate average energies are obtained, of course, by considering replica symmetry breaking, for instance the 1RSB equations considered in sec.1.4, and working with a probability distribution of the BP fixed points.

### 3.2.1 Quantum ising model

We consider the quantum Ising model with Hamiltonian  $H = H_0 + H_1$  where  $H_0 = -\sum_{(ij) \in \mathcal{E}_q} J_{ij} \sigma_i^z \sigma_j^z$  and  $H_1 = -\sum_{i=1}^N h_i \sigma_i^x$ . The  $J_{ij}$  are real parameter describing the interactions in the  $z$  direction between spins, and the  $h_i$  are also real parameter representing transverse external fields on the  $x$  direction. The interaction graph is defined by  $\mathcal{E}_q$ , and  $N$  is the size of the system. The  $\sigma^{x,y,z}$  are the standard Pauli matrices. The states  $|\boldsymbol{\sigma}\rangle$  are the  $2^N$  configurations of the  $\sigma^z$  spins. In this case  $\langle \boldsymbol{\sigma} | H_0 | \boldsymbol{\sigma} \rangle = -\sum_{(ij) \in \mathcal{E}_q} J_{ij} \sigma_i \sigma_j$  and  $\langle \boldsymbol{\sigma}' | H_1 | \boldsymbol{\sigma} \rangle = -\sum_i h_i \delta_{\sigma_i, -\sigma'_i} \delta_{\boldsymbol{\sigma} \setminus i, \boldsymbol{\sigma}' \setminus i}$ , where  $\delta_{a,b}$  is the the delta of Dirac between two variable  $(a,b)$ . We consider only real parameter in the Hamiltonian, then the ground state wave function can be always describe by a real function. Moreover, if  $h > 0$ , the ground state is a positive wave function<sup>1</sup>. In this thesis we consider only positive  $h$  values of transverse external field.

---

<sup>1</sup>giustify

### 3.2.2 Pairwise model

For the trial wave functions we take a pairwise model plus external field ansatz:

$$\psi(\boldsymbol{\sigma}; \mathbf{P}) = \frac{e^{\sum_i B_i \sigma_i + \sum_{(ij) \in \mathcal{E}_c} K_{ij} \sigma_i \sigma_j}}{\left( \sum_{\boldsymbol{\sigma}} e^{\sum_i 2B_i \sigma_i + \sum_{(ij) \in \mathcal{E}_c} 2K_{ij} \sigma_i \sigma_j} \right)^{1/2}}, \quad (3.8)$$

with real parameters  $\mathbf{P} = \{B_i, K_{ij} | i = 1, \dots, N, (ij) \in \mathcal{E}_c\}$ . This results in the Gibbs measure  $\mu(\boldsymbol{\sigma}; \mathbf{P}) = |a(\boldsymbol{\sigma}; \mathbf{P})|^2$  of a classical spin-glass model with external fields  $2B_i$  and couplings  $2K_{ij}$ . Notice that the classical interaction graph  $\mathcal{E}_c$  could be different from the quantum one  $\mathcal{E}_q$ . For simplicity, in the following we will assume that the two coincide as happens in zero transverse fields; better representations could be obtained by adding the higher order neighbors to  $\mathcal{E}_c$ .

Given the classical measure, the BP equations read:

$$\mu_{i \rightarrow j}(\sigma_i) \propto e^{2B_i \sigma_i} \prod_{k \in \partial i \setminus j} \left( \sum_{\sigma_k} e^{2K_{ik} \sigma_i \sigma_k} \mu_{k \rightarrow i}(\sigma_k) \right) \equiv \mathcal{BP}_{i \rightarrow j}, \quad (3.9)$$

where  $\partial i$  refers to the set of spins interacting with spin  $i$  in  $\mathcal{E}_c$ . Figure 3.1 displays the set of variables and interactions in the classical interaction graph. Having the cavity marginals, the average of local energies  $e_{ij} \equiv -J_{ij} \sigma_i^z \sigma_j^z$  and  $e_i \equiv -h_i \sigma_i^x$  read

$$\langle e_{ij} \rangle_\mu = -J_{ij} \sum_{\sigma_i, \sigma_j} \sigma_i \sigma_j \mu_{ij}(\sigma_i, \sigma_j), \quad (3.10)$$

$$\langle e_i \rangle_\mu = -h_i \sum_{\sigma_i, \sigma_{\partial i}} e^{-2B_i \sigma_i - \sum_{j \in \partial i} 2K_{ij} \sigma_i \sigma_j} \mu_{i, \partial i}(\sigma_i, \sigma_{\partial i}), \quad (3.11)$$

with the local BP marginals  $\mu_{ij}(\sigma_i, \sigma_j) \propto e^{2K_{ij} \sigma_i \sigma_j} \mu_{i \rightarrow j}(\sigma_i) \mu_{j \rightarrow i}(\sigma_j)$  and  $\mu_{i, \partial i}(\sigma_i, \sigma_{\partial i}) \propto e^{2B_i \sigma_i} \prod_{j \in \partial i} [e^{2K_{ij} \sigma_i \sigma_j} \mu_{j \rightarrow i}(\sigma_j)]$ .

The above average energies define the Boltzmann weight  $e^{-\beta_{opt} \langle E_0 \rangle_\mu - \beta_{opt} \langle E_1 \rangle_\mu}$  for a given configuration of the variational parameters and the BP messages; see figure 3.1. The cavity marginals of the parameters (including the BP messages) can be written in a higher-level Bethe approximation, resembling the one-step RSB equations (sec.[1.4]):

$$M_{i \rightarrow j}(K_{ij}, \mu_{ij}) \propto e^{-\beta_{opt} \langle e_{ij} \rangle_\mu} \sum_{\substack{B_i \\ \{K_{ik}, \mu_{ik} | \\ k \in \partial i \setminus j\}}} e^{-\beta_{opt} \langle e_i \rangle_\mu} \prod_{k \in \partial i \setminus j} M_{k \rightarrow i}(K_{ik}, \mu_{ik}) \mathbb{1}_{BP(i)}, \quad (3.12)$$

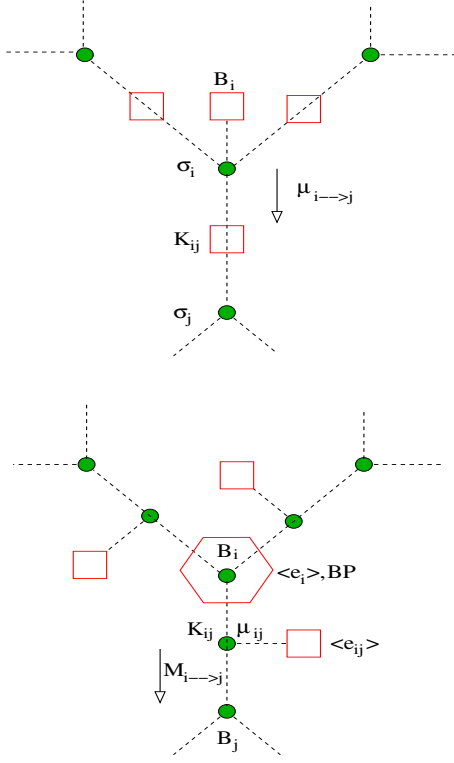


FIGURE 3.1: (top) The interaction graphs of the classical system representing the trial wave function and (bottom) the resulting variational problem. The variables are shown with solid circles, and the interactions are shown with open polygons.

where for brevity we defined  $\mu_{ij} \equiv \{\mu_{i \rightarrow j}, \mu_{j \rightarrow i}\}$  and  $\mathbb{1}_{BP(i)} \equiv \prod_{j \in \partial i} \delta(\mu_{i \rightarrow j} - \mathcal{BP}_{i \rightarrow j})$ . The Bethe free energy is given by  $F = \sum_i F_i - \sum_{(ij) \in \mathcal{E}_c} F_{ij}$  with

$$e^{-\beta_{opt} F_i} \equiv \sum_{B_i, \{K_{ij}, \mu_{ij} | j \in \partial i\}} e^{-\beta_{opt} \langle e_i \rangle_\mu} \prod_{j \in \partial i} M_{j \rightarrow i}(K_{ij}, \mu_{ij}) \mathbb{1}_{BP(i)}, \quad (3.13)$$

$$e^{-\beta_{opt} F_{ij}} \equiv \sum_{K_{ij}, \mu_{ij}} e^{+\beta_{opt} \langle e_{ij} \rangle_\mu} M_{i \rightarrow j}(K_{ij}, \mu_{ij}) M_{j \rightarrow i}(K_{ij}, \mu_{ij}), \quad (3.14)$$

Notice that in the last equation we have the positive sign in the exponential to count correctly the energy contribution  $\langle e_{ij} \rangle_\mu$ .

The  $\beta_{opt} \rightarrow \infty$  limit of the above equations, taking the scaling  $M_{i \rightarrow j}(K_{ij}, \mu_{ij}) = e^{\beta_{opt} \mathcal{M}_{i \rightarrow j}(K_{ij}, \mu_{ij})}$ , read

$$\mathcal{M}_{i \rightarrow j}(K_{ij}, \mu_{ij}) = -\langle e_{ij} \rangle_\mu + \max_{B_i, \{K_{ik}, \mu_{ik} | k \in \partial i \setminus j\}: \mathbb{1}_{BP(i)}} \left\{ -\langle e_i \rangle_\mu + \sum_{k \in \partial i \setminus j} \mathcal{M}_{k \rightarrow i}(K_{ik}, \mu_{ik}) \right\}, \quad (3.15)$$



they are the Max-Sum/BP equations like in the stochastic optimization (sec.[1.4.1]). The minimum energy is given by  $E_g = \lim_{\beta_{opt} \rightarrow \infty} F = -(\sum_i e_i - \sum_{(ij) \in \mathcal{E}_c} e_{ij})$  with the local energy shifts

$$e_i \equiv \max_{B_i, \{K_{ij}, \mu_{ij} | j \in \partial i\}: \mathbb{1}_{BP(i)}} \left\{ -\langle e_i \rangle_\mu + \sum_{j \in \partial i} \mathcal{M}_{j \rightarrow i}(K_{ij}, \mu_{ij}) \right\}, \quad (3.16)$$

$$e_{ij} \equiv \max_{K_{ij}, \mu_{ij}} \{ \langle e_{ij} \rangle_\mu + \mathcal{M}_{i \rightarrow j}(K_{ij}, \mu_{ij}) + \mathcal{M}_{j \rightarrow i}(K_{ij}, \mu_{ij}) \}. \quad (3.17)$$

Before solving the above equations we will consider some simpler cases.

### 3.2.3 Zero couplings: Mean field solution

In the zeroth order of the approximation we take  $K_{ij} = 0$  for any edge  $(ij)$ . This is a mean-field approximation with a factorized measure  $\mu(\underline{\sigma}) = \prod_i \mu_i(\sigma_i)$ , where  $\mu_i(\sigma_i) = e^{2B_i \sigma_i} / [2 \cosh(2B_i)]$ . Then using equations (3.10) and (3.11) we obtain the average local energies:  $\langle e_{ij} \rangle_\mu = -J_{ij} \tanh(2B_i) \tanh(2B_j)$  and  $\langle e_i \rangle_\mu = -h_i / \cosh(2B_i)$ ; therefore

$$\langle \psi(\underline{B}) | H | \psi(\underline{B}) \rangle = - \sum_{(ij) \in \mathcal{E}_q} J_{ij} \tanh(2B_i) \tanh(2B_j) - \sum_i \frac{h_i}{\cosh(2B_i)}. \quad (3.18)$$

Here we write directly the MaxSum equations that can be used to estimate the optimal parameters and the minimum average energy:

$$\mathcal{M}_{i \rightarrow j}(B_i) = -\langle e_i \rangle_\mu + \sum_{k \in \partial i \setminus j} \max_{B_k} \{ -\langle e_{ik} \rangle_\mu + \mathcal{M}_{k \rightarrow i}(B_k) \}. \quad (3.19)$$

Then we find the optimal paramters by maximizing the local MaxSum weights:

$$B_i^* = \arg \max_{B_i} \left\{ -\langle e_i \rangle_\mu + \sum_{j \in \partial i} \max_{B_j} \{ -\langle e_{ij} \rangle_\mu + \mathcal{M}_{j \rightarrow i}(B_j) \} \right\}. \quad (3.20)$$

### 3.2.4 Zero fields: Symmetric solution

As long as the fields  $B_i$  are zero we have always a symmetric solution  $\mu_{i \rightarrow j}(\sigma_i) = 1/2$  to the BP equations in the classical system. This gives the average local

energies:  $\langle e_{ij} \rangle_\mu = -J_{ij} \tanh(2K_{ij})$  and  $\langle e_i \rangle_\mu = -h_i / [\prod_{j \in \partial i} \cosh(2K_{ij})]$ , and

$$\langle \psi(\underline{K}) | H | \psi(\underline{K}) \rangle = - \sum_{(ij) \in \mathcal{E}_q} J_{ij} \tanh(2K_{ij}) - \sum_i \frac{h_i}{\prod_{j \in \partial i} \cosh(2K_{ij})}. \quad (3.21)$$

The resulting MaxSum equations are

$$\mathcal{M}_{i \rightarrow j}(K_{ij}) = -\langle e_{ij} \rangle_\mu + \max_{\{K_{ik} | k \in \partial i \setminus j\}} \left\{ -\langle e_i \rangle_\mu + \sum_{k \in \partial i \setminus j} \mathcal{M}_{k \rightarrow i}(K_{ik}) \right\}, \quad (3.22)$$

and the optimal couplings are estimated by

$$K_{ij}^* = \arg \max_{K_{ij}} \{ \langle e_{ij} \rangle_\mu + \mathcal{M}_{i \rightarrow j}(K_{ij}) + \mathcal{M}_{j \rightarrow i}(K_{ij}) \}. \quad (3.23)$$

Notice that here we have  $\langle \sigma_i^z \rangle = 0$ , which is not the case in the ordered phase. In addition, the symmetric solution does not give an accurate average energy when replica symmetry is broken, which may happen for large couplings in the classical system; the Bethe approximation works well when distant spins are nearly independent, whereas the symmetric solution does not respect this property in an RSB phase. To get around this problem one can work with the nontrivial BP fixed points, e.g., by demanding a total magnetization of magnitude greater than  $\delta m \ll 1$ . At the same time one may need to limit the range of couplings to  $|K_{ij}| < K_{max}$  in order to avoid dominance by very large couplings.

### 3.2.5 General solution

In general to solve the MaxSum-BP equations we have to work with discrete fields  $B_l \in \{l\delta B | l = -L_B, \dots, L_B\}$ , couplings  $K_l \in \{l\delta K | l = -L_K, \dots, L_K\}$ , and BP cavity fields  $\nu_l \in \{l\delta \nu | l = -L_\nu, \dots, L_\nu\}$ . The BP cavity fields  $\nu_{i \rightarrow j}$  are defined by  $\mu_{i \rightarrow j}(\sigma_i) \propto e^{\nu_{i \rightarrow j} \sigma_i}$ . An exhaustive solution of the MaxSum-BP equations would take a time of order  $Nd(2L_B)[(2L_K)(2L_\nu)]^d$  where  $d$  is the maximum degree in  $\mathcal{E}_c$ . Notice that given the couplings  $K_{ij}$  and the input BP messages  $\nu_{j \rightarrow i}$  around spin  $i$ , one obtains  $\nu_{i \rightarrow j}$  for any value of  $B_i$  from the BP equations. The above equations can be solved more efficiently (for large degrees) by using a convolution function of four variables (needed to compute  $\langle e_i \rangle_\mu$  in the MaxSum-BP equations) resulting in a time complexity of order  $Nd(2L_K)(2L_\nu)^6$ ; see appendix in the article[. In the following instead we use a computationally easier but approximate way of

solving the equations by restricting the domain of variables: We start by assigning a small number of randomly selected states  $S_{ij} = \{(K_{ij}^1, \nu_{ij}^1), \dots, (K_{ij}^S, \nu_{ij}^S)\}$  to each variable  $(K_{ij}, \nu_{ij})$ . Then we run the MaxSum-BP equations with these restricted search spaces to converge the equations and sort the states in  $S_{ij}$  according to their MaxSum-BP weights:

$$w_{ij}(K_{ij}^l, \nu_{ij}^l) = \langle e_{ij} \rangle_\mu + \mathcal{M}_{i \rightarrow j}(K_{ij}^l, \nu_{ij}^l) + \mathcal{M}_{j \rightarrow i}(K_{ij}^l, \nu_{ij}^l). \quad (3.24)$$

Next we update the search spaces by replacing the states having smaller weights with some other ones generated randomly but close to the best state observed during the algorithm. The above two steps are repeated to find better search spaces and therefore parameters. The algorithm performance would depend on the size of the search spaces, approaching the correct one for  $S \rightarrow \infty$ .

The search spaces at the beginning are chosen randomly therefore we introduce a tolerance  $\delta\nu$  to accept those BP messages that satisfy the BP equations within  $\pm\delta\nu$ . One may start from a large tolerance and decrease it slowly after each update of the search spaces. The mean-field solution mentioned before can provide a good initial point for the search spaces.

### 3.2.6 Numerical results

#### 3.2.6.1 Random coupling chain

First, we consider a spin chain in uniform and positive transverse field but with random couplings. In figure 3.2 we compare the results obtained by the above approximations with the exact ones obtained by the modified Lanczos method [66]. As expected, the mean-field ansatz is better than the symmetric solution in the ordered phase where the local magnetizations are nonzero. The reverse happens in the disordered phase and even in the ordered phase close to the transition point where the local magnetizations are still small. The two limiting behaviors are therefore displayed in the general solution.

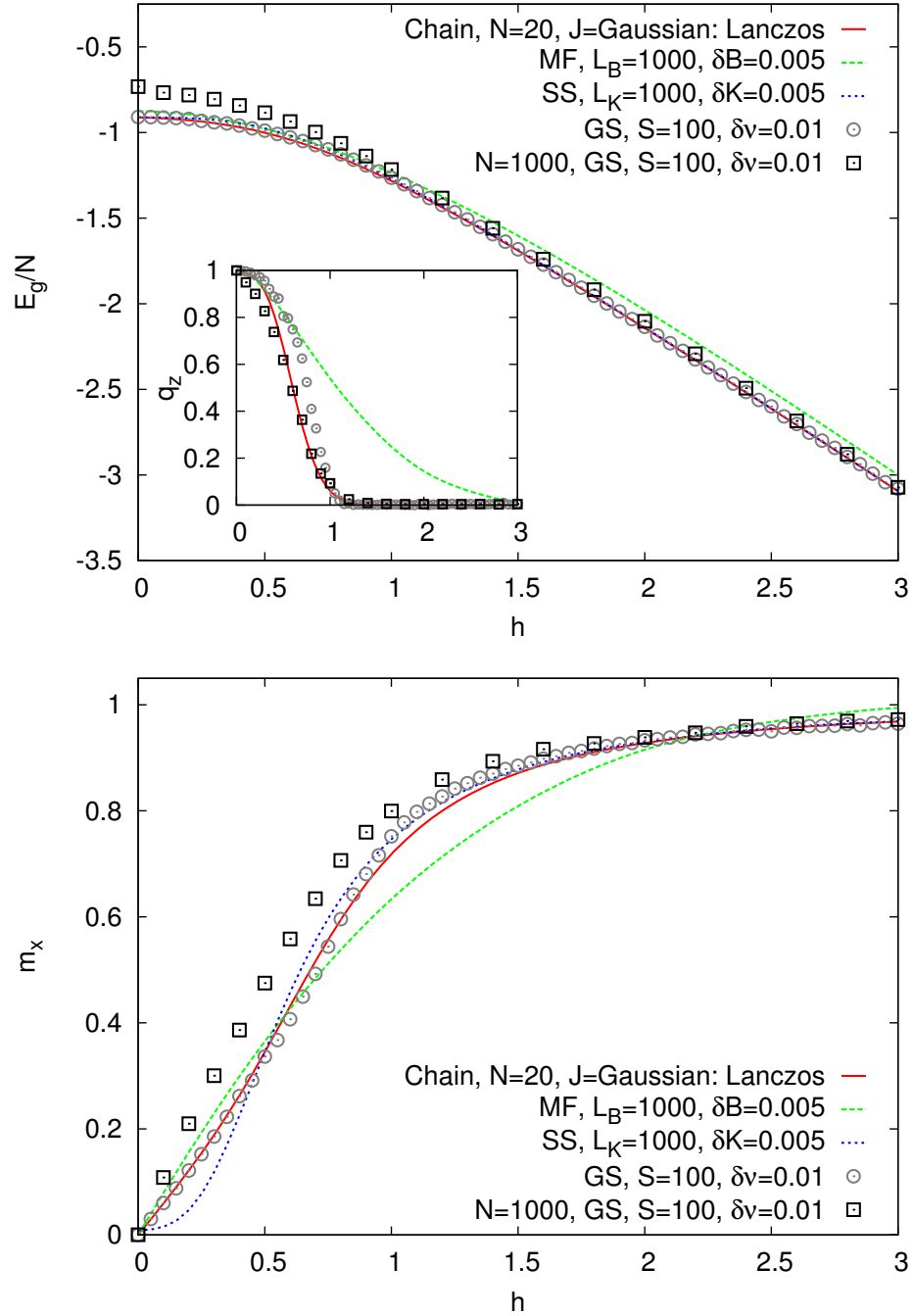


FIGURE 3.2: Comparing the exact results with those of the mean-field (MF), the symmetric solution (SS) and the general solution (GS) in a single instance of the quantum Ising model in transverse field  $h$  on a chain of size  $N = 20$ . For comparison we also display the GS results for  $N = 1000$ . Here  $E_g$  and  $m_x$  are the ground-state energy and magnetization in the  $x$  direction, respectively. The inset shows the Edwards-Anderson order parameter  $q_z = (\sum_i \langle \sigma_i^z \rangle^2)/N$ . The couplings  $J_{ij}$  are Gaussian random numbers of mean zero and variance one.  $\delta B$ ,  $\delta K$ , and  $\delta \nu$  are the sizes of bins in the discrete representation of the parameters, and  $S$  is the number of states in the restricted domains. The data for GS are obtained by restricting the search algorithm to total magnetizations of magnitude greater than  $\delta m = 0.05$ .

### 3.2.6.2 Random graph

As another example we study the same model on a single instance of random regular graphs where each spin interacts with a fixed number of other randomly selected spins. The results displayed in figure 3.3 show similar behaviors observed above, except that in the ordered phase the symmetric solution gives a lower ground state energy than the exact one. As explained before, this is due to the poor estimation of the average energy by the symmetric solution when there are multiple BP fixed points. Moreover, due to the loops and small size of the system we find larger deviations from the exact data compared to the chain model.

Similar qualitative behaviors are observed in ferromagnetic ( $J_{ij} = 1$ ) and  $\pm J$  spin-glass ( $J_{ij} = \pm 1$  with equal probability) models on a random regular graph of degree  $d = 3$ . In the ferromagnetic case, a reasonable trial wave function is obtained by taking  $B_i = B$  and  $K_{ij} = K$ . This simplification allows us to find easily the optimal parameters, and so the critical field  $h_c^{ferro} \sim 2.29$  in the thermodynamic limit. Figure 3.4 displays the phase diagram of the ferromagnetic model obtained in this way. For the  $\pm J$  spin-glass model, the general solution on single instances of size  $N = 1000$  gives  $h_c^{\pm J} \simeq 2.0$ . The corresponding values in the thermodynamic limit given in Refs. [13] and [12] are:  $h_c^{ferro} \sim 2.23$  and  $h_c^{\pm J} \sim 1.77$ .

## 3.3 Low temperature excitations

In this section we use the variational quantum cavity method, shown in the previous section, to find approximate solutions for the excited-states of the transverse Ising model. The main point of this study is to write the overlap between two wave functions in the Bethe approximation, which is asymptotically exact as long as the trial wave functions can be represented by classical systems of locally tree-like interaction graphs. This allows us to replace the global orthogonality condition with some local constraints on the variational parameters and the cavity messages in the Bethe expression for the overlap. In summary, given an appropriate trial wave function, we evaluate the Hamiltonian expectation in the subspace of wave functions orthogonal to the lower-energy states of the Hamiltonian and find an estimation of the optimal variational parameters, all within the Bethe approximation. We implemented it by a local message-passing algorithm.

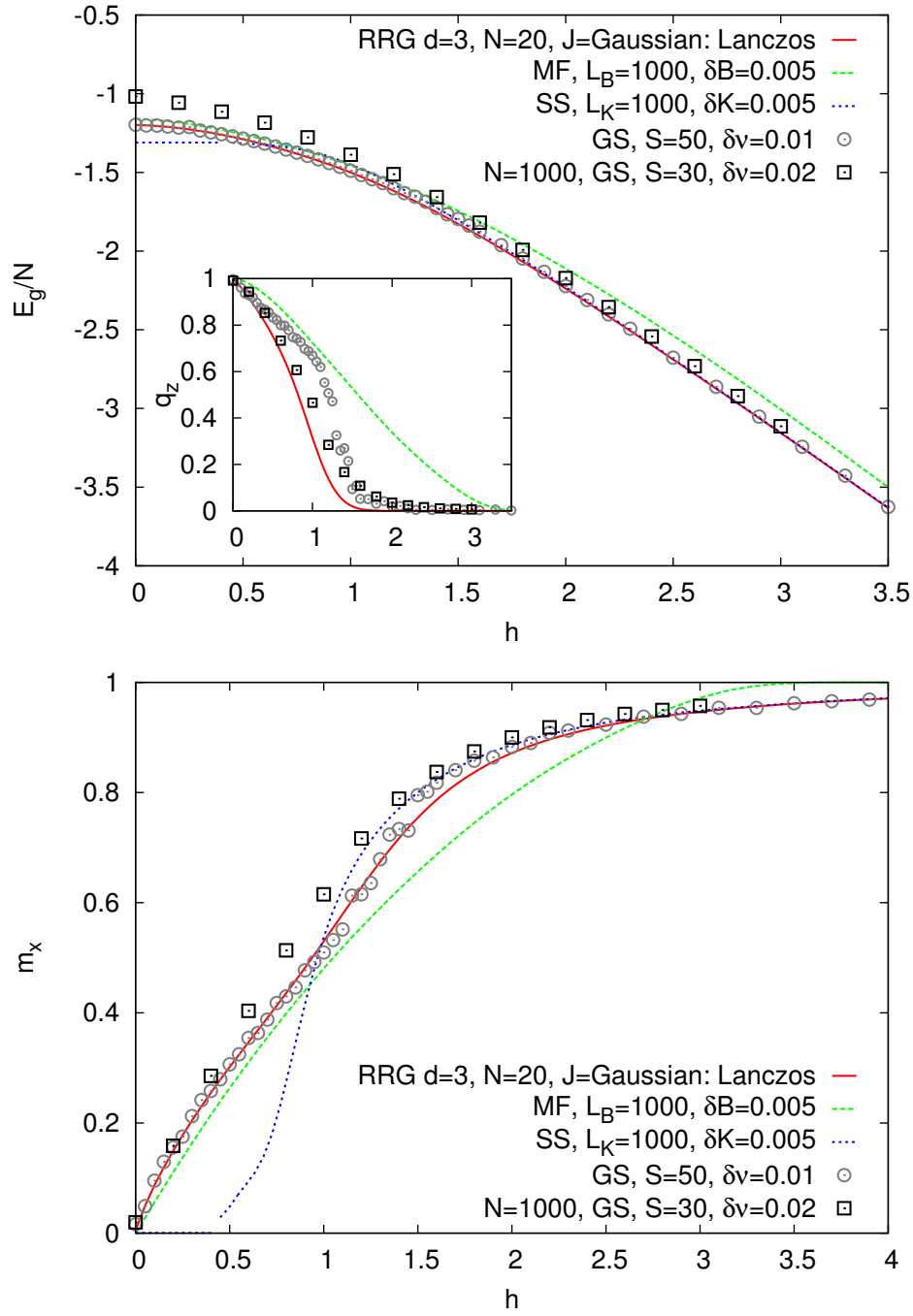


FIGURE 3.3: The quantum Ising model in transverse field  $h$  on a random regular graph (RRG) of degree  $d = 3$  and size  $N = 20$ . For comparison we also display the GS results for  $N = 1000$ . The couplings  $J_{ij}$  are Gaussian random numbers of mean zero and variance one. The data for GS are obtained by restricting the search algorithm to total magnetizations of magnitude greater than  $\delta m = 0.05$  and couplings of magnitude less than  $K_{max} = 1$ .

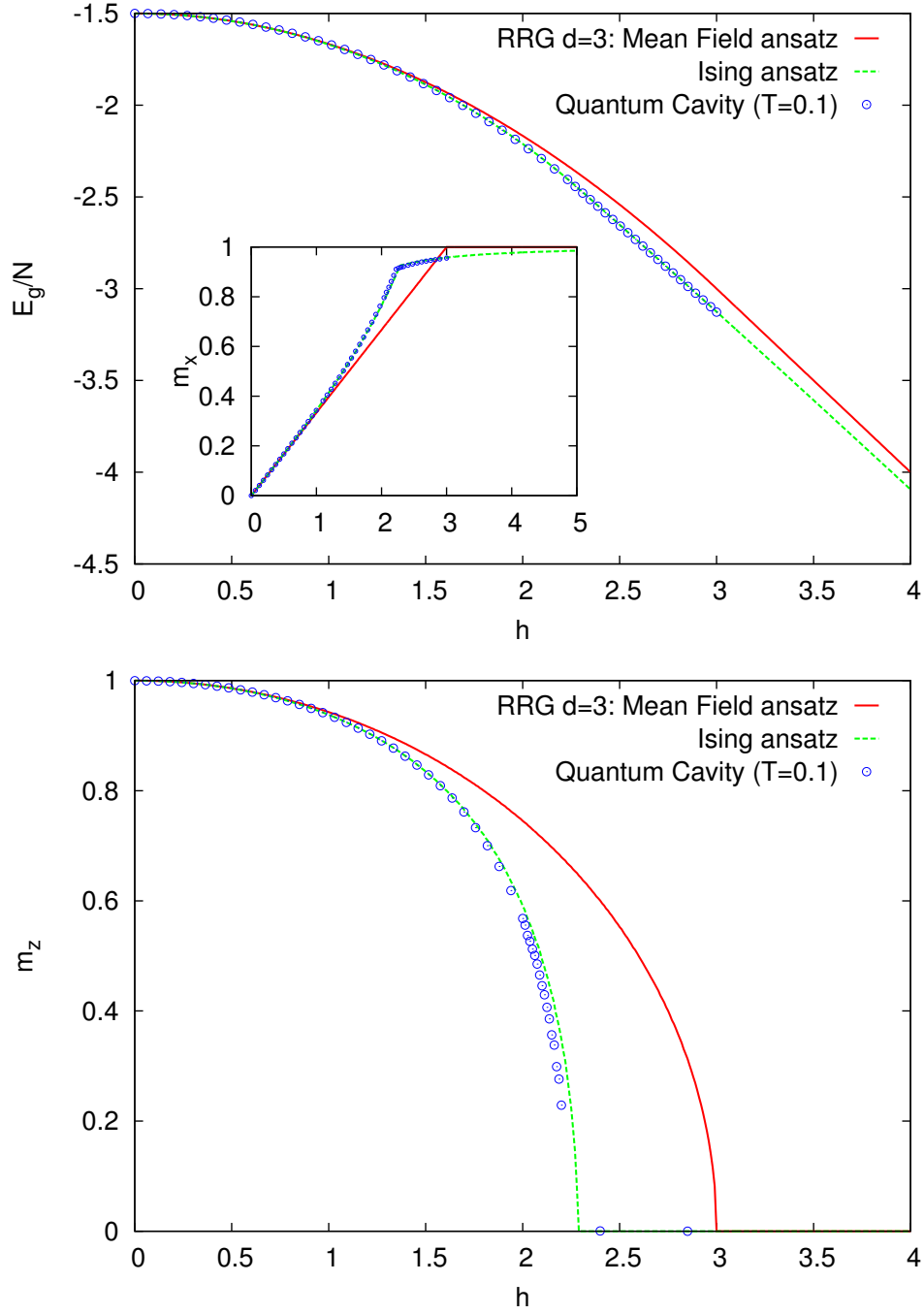


FIGURE 3.4: The quantum Ising model with ferromagnetic interactions ( $J_{ij} = 1$ ) in transverse field  $h$  on RRG of degree  $d = 3$ . The results of homogeneous mean field  $B_i = B$  and Ising ( $B_i = B, K_{ij} = K$ ) trial wave functions are compared with the results of the quantum cavity method [13] for a small temperature in the thermodynamic limit.

Note that, usually, for large and disordered systems we do not have the exact ground-state of the system. The best that we can do is to find a good variational wave function that represent the low-energy state of the system. Obviously, an excited state that is orthogonal to an approximate ground state does not necessarily provide an upper bound for the excited-state energy. There are, of course, other ways to find the excited states that do not rely on the orthogonality to a priori known ground state, but we found the above procedure more amenable to the cavity method that we are going to use in this paper. Nevertheless, any orthogonal set of quantum states defines a subspace of wave vectors that according to the Courant min-max theorem [67] can be used to find some upper bounds for the Hamiltonian eigenvalues.

### 3.3.1 Orthogonality constraints

Let us denote by  $|\Psi_n(\underline{P}^n)\rangle$  the  $n$ th trial wave function that minimizes the average energy  $\langle \Psi_n(\underline{P}^n) | H | \Psi_n(\underline{P}^n) \rangle$  conditioned on the orthogonality constraints  $\langle \Psi_n(\underline{P}^n) | \Psi_m(\underline{P}^m) \rangle = 0$  for  $m = 0, \dots, n-1$ . The corresponding classical systems are represented by measures  $\mu_n(\underline{\sigma}; \underline{P}^n)$ . In the following we are going to satisfy the orthogonality constraints within the Bethe approximation,

$$\langle \Psi_n(\underline{P}^n) | \Psi_m(\underline{P}^m) \rangle = \sum_{\underline{\sigma}} \psi_n^*(\underline{\sigma}; \underline{P}^n) \psi_m(\underline{\sigma}; \underline{P}^m) \simeq e^{-(\sum_i F_i + \sum_a F_a - \sum_{(ia)} F_{ia})} = \emptyset \quad (3.25)$$

where  $F_i$ ,  $F_a$ , and  $F_{ia}$  are the free energy changes by adding variable node  $i$ , interaction node  $a$ , and link  $(ia)$  to the complex measure  $\nu_{n,m}(\underline{\sigma}; \underline{P}^n, \underline{P}^m) \propto \psi_n^*(\underline{\sigma}; \underline{P}^n) \psi_m(\underline{\sigma}; \underline{P}^m) \propto \prod_a \phi_a^*(\underline{\sigma}^{\partial a}; P_a^n) \phi_a(\underline{\sigma}^{\partial a}; P_a^m)$ . These quantities are given by (see sec.1.3.2),

$$e^{-F_i} = \sum_{\sigma_i} \prod_{a \in \partial i} \nu_{a \rightarrow i}(\sigma_i), \quad (3.26)$$

$$e^{-F_a} = \sum_{\underline{\sigma}^{\partial a}} \phi_a^*(\underline{\sigma}^{\partial a}; P_a^n) \phi_a(\underline{\sigma}^{\partial a}; P_a^m) \prod_{i \in \partial a} \nu_{i \rightarrow a}(\sigma_i) \quad (3.27)$$

$$e^{-F_{ia}} = \sum_{\sigma_i} \nu_{i \rightarrow a}(\sigma_i) \nu_{a \rightarrow i}(\sigma_i), \quad (3.28)$$

where the cavity marginals  $\nu_{i \rightarrow a}(\sigma_i)$  and  $\nu_{a \rightarrow i}(\sigma_i)$  satisfy the BP equations for the complex measure  $\nu_{n,m}(\underline{\sigma}; \underline{P}^n, \underline{P}^m)$ . Thus, to have orthogonality it is enough to



have  $e^{-F_a} = 0$  for some  $a$ . This defines a constraint on the parameter  $P_a^n$  given  $\underline{P}^n$ .

In summary, to estimate the average energy and to satisfy the orthogonality constraints we need to know the BP marginals of the classical measure  $\mu_n$  and  $\{\nu_{n,m}|m = 0, \dots, n-1\}$ . In addition, we have to choose the set of constrained parameters  $A_n \equiv \{a_m|m = 0, \dots, n-1\}$  for the orthogonality constraints  $e^{-F_{a_m}} = 0$ . In this paper we will use a greedy strategy to construct  $A_n$ , by choosing the parameters that at least locally minimize the energy expectation. Finally, the problem of minimizing over the variational parameters is:

$$\mathcal{Z}_n = \sum_{\underline{P}^n} \sum_{\mu_n, \{\nu_{n,m}\}} \mathbb{I}_{BP} \prod_{m=0, \dots, n-1} \mathbb{I}_{n,m} e^{-\beta_{opt} \langle E(\underline{\sigma}) \rangle_{\mu_n}}, \quad (3.29)$$

where for  $\beta_{opt} \rightarrow \infty$  the Gibbs measure is concentrated on the optimal parameters. The indicator functions  $\mathbb{I}_{BP}$  and  $\mathbb{I}_{n,m}$  ensure that the messages  $\mu_n, \{\nu_{n,m}\}$  satisfy the BP equations and the states  $n$  and  $m$  are orthogonal. Starting from  $n = 0$ , one can find the other states one by one after solving the above optimization problem.

One may find an approximate solution to the above problem by a two-stage algorithm: Given  $\{\underline{P}^m|m = 0, \dots, n-1\}$  and an arbitrary set of the constrained parameters  $A_n$ , we run BP to find the set of marginals  $\{\nu_{n,m}|m = 0, \dots, n-1\}$ . These are used to fix the constrained parameters in  $A_n$  to satisfy the orthogonality constraints. Then we minimize the average energy  $\langle E(\underline{\sigma}) \rangle_{\mu_n}$  over the remaining parameters. The above two stages are repeated to converge the algorithm.

### 3.3.2 The mean-field approximation

Let us start with the mean-field (MF) approximation, where the trial wave functions are represented by the product states:

$$\psi(\underline{\sigma}; \underline{B}) \propto \prod_i e^{B_i \sigma_i}, \quad (3.30)$$

with complex parameters  $B_i$ . This results to  $e_i(\sigma_i) = -h_i e^{-2B_i^R \sigma_i} \cos(2B_i^I \sigma_i)$  and the following classical measure  $\mu(\underline{\sigma}; \underline{B}) \propto \prod_i e^{2B_i^R \sigma_i}$ . By superscripts  $R$  and  $I$  we mean the real and imaginary part of the parameters. Given the above measure we

find

$$\langle e_i(\sigma_i) \rangle_\mu = -h_i \frac{\cos(2B_i^I)}{\cosh(2B_i^R)}, \quad (3.31)$$

$$\langle e_{ij}(\sigma_i, \sigma_j) \rangle_\mu = -J_{ij} \tanh(2B_i^R) \tanh(2B_j^R). \quad (3.32)$$

We see that for non-negative  $h_i$  the average energy is minimized by setting  $B_i^I = 0$ . Therefore, as long as we are interested in the ground state, we can set the imaginary parts to zero. In this case, the variational problem reads

$$\mathcal{Z}_0 = \sum_{\underline{B}} e^{-\beta_{opt} \sum_{(ij) \in \mathcal{E}_q} \langle e_{ij} \rangle_\mu - \beta_{opt} \sum_i \langle e_i \rangle_\mu}, \quad (3.33)$$

The cavity marginals of the parameters in the Bethe approximation are

$$M_{i \rightarrow j}(B_i) \propto e^{-\beta_{opt} \langle e_i \rangle_\mu} \prod_{k \in \partial i \setminus j} \left( \sum_{B_k} e^{-\beta_{opt} \langle e_{ik} \rangle_\mu} M_{k \rightarrow i}(B_k) \right). \quad (3.34)$$

For  $\beta_{opt} \rightarrow \infty$  and scaling  $M_{i \rightarrow j}(B_i) = e^{-\beta_{opt} \mathcal{M}_{i \rightarrow j}(B_i)}$  we find the Min-Sum equations [68]:

$$\mathcal{M}_{i \rightarrow j}(B_i) = \langle e_i \rangle_\mu + \sum_{k \in \partial i \setminus j} \min_{B_k} \{ \langle e_{ik} \rangle_\mu + \mathcal{M}_{k \rightarrow i}(B_k) \}. \quad (3.35)$$

The equations are solved by iteration starting from random initial messages. After each iteration we subtract a constant from the messages to have  $\min_{B_i} \mathcal{M}_{i \rightarrow j}(B_i) = 0$ . Then we find the optimal parameters by minimizing the local Min-Sum weights,

$$B_i^0 = \arg \min_{B_i} \left\{ \langle e_i \rangle_\mu + \sum_{j \in \partial i} \min_{B_j} \{ \langle e_{ij} \rangle_\mu + \mathcal{M}_{j \rightarrow i}(B_j) \} \right\}. \quad (3.36)$$

Note that in the MF approximation the orthogonality condition reads

$$\langle \Psi_n(\underline{B}^n) | \Psi_m(\underline{B}^m) \rangle \propto \prod_i \cosh(B_i^{n*} + B_i^m) = 0, \quad (3.37)$$

thus, it is enough to have  $\cosh(B_i^{n*} + B_i^m) = 0$  for some  $i$ . This means  $B_i^{nR} = -B_i^{mR}$  and  $B_i^{nI} - B_i^{mI} = \pi/2$ . Therefore, the  $n$ th excited state can be obtained by the

following set of constraints:

$$\mathbb{I}_{n,m} = \begin{cases} B_{i_m}^{nR} + B_{i_m}^{mR} = 0, \\ B_{i_m}^{nI} - B_{i_m}^{mI} = \frac{\pi}{2}, \end{cases} \quad (3.38)$$

for  $m = 0, \dots, n-1$ . The index  $i_m$  is chosen to minimize the local energy  $\langle e_{i_m}(\sigma_i) \rangle_{\mu_n}$ . Then, one can use the same Min-Sum equations as above to minimize over the remaining parameters. In this way we can find at most  $N$  orthogonal product states of minimum energies  $E_n$ .

Given the states  $|\Psi_m(\underline{B}^m)\rangle$  for  $m = 0, \dots, n$ , we can easily compute the Hamiltonian matrix elements  $H_{mm'} \equiv \langle \Psi_m(\underline{B}^m) | H | \Psi_{m'}(\underline{B}^{m'}) \rangle$ , which for arbitrary parameters read

$$H_{mm'} = \prod_i \left( \frac{\cosh(B_i^{m*} + B_i^{m'})}{(\cosh(2B_i^{mR}) \cosh(2B_i^{m'R}))^{1/2}} \right) \times \left\{ - \sum_{(ij) \in \mathcal{E}_q} J_{ij} \tanh(B_i^{m*} + B_i^{m'}) \tanh(B_j^{m*} + B_j^{m'}) - \sum_i h_i \frac{\cosh(B_i^{m*} - B_i^{m'})}{\cosh(B_i^{m*} + B_i^{m'})} \right\}. \quad (3.39)$$

One can diagonalize the Hamiltonian in the subspace spanned by the above states to obtain the eigenvalues  $\lambda_m$ . Then, using the min-max principle, we know that  $\lambda_n$  is an upper bound for the  $n$ th eigenvalue of the Hamiltonian. Indeed, we found that in this case  $\lambda_n \simeq E_n$  for large  $N$  as the off-diagonal matrix elements  $H_{mm'}$  decay exponentially with the size of system.

Notice that instead of imposing the orthogonality exactly we could ask for an exponentially small overlap  $\langle \Psi_n(\underline{B}^n) | \Psi_m(\underline{B}^m) \rangle < \epsilon^N$  by demanding

$$\frac{\cosh(B_i^{n*} + B_i^m)}{(\cosh(2B_i^{nR}) \cosh(2B_i^{mR}))^{1/2}} < \epsilon \leq 1, \quad (3.40)$$

for all  $i$ , which can easily be imposed in the above Min-Sum equations. Indeed to apply the min-max theorem we do not need a set of orthogonal states [67]; according to the theorem, the  $n$ th eigenvalue is given by  $\min_{S_{n+1}} \max_{|\psi\rangle \in S_{n+1}: \langle \psi | \psi \rangle = 1} \langle \psi | H | \psi \rangle$  where  $S_{n+1}$  is any subspace of dimension  $n+1$ .

### 3.3.3 Beyond the mean-field approximation

We can do better than the MF approximation by adding the local two-body or Jastrow interactions [69] to the trial wave functions:

$$\psi(\underline{\sigma}; \underline{P}) \propto \prod_i \phi_i(\sigma_i) \prod_{(ij) \in \mathcal{E}_c} \phi_{ij}(\sigma_i, \sigma_j), \quad (3.41)$$

with

$$\phi_i(\sigma_i) \equiv e^{B_i \sigma_i}, \quad \phi_{ij}(\sigma_i, \sigma_j) \equiv e^{K_{ij} \sigma_i \sigma_j}. \quad (3.42)$$

For simplicity we are going to assume  $\mathcal{E}_c = \mathcal{E}_q$ . As a result, we obtain

$$e_i(\sigma_i, \underline{\sigma}^{\partial i}) = -h_i e^{-2B_i^R \sigma_i - \sum_{j \in \partial i} 2K_{ij}^R \sigma_i \sigma_j} \cos(2B_i^I \sigma_i + \sum_{j \in \partial i} 2K_{ij}^I \sigma_i \sigma_j). \quad (3.43)$$

The average energy is computed with respect to the following classical measure  $\mu(\underline{\sigma}) \propto \prod_i |\phi_i(\sigma_i)|^2 \prod_{(ij) \in \mathcal{E}_c} |\phi_{ij}(\sigma_i, \sigma_j)|^2$ . To estimate the average energies  $\langle e_{ij}(\sigma_i, \sigma_j) \rangle_\mu$  and  $\langle e_i(\sigma_i, \underline{\sigma}^{\partial i}) \rangle_\mu$  we need the following local marginals

$$\mu_{ij}(\sigma_i, \sigma_j) = \frac{1}{Z_{ij}} |\phi_{ij}(\sigma_i, \sigma_j)|^2 \mu_{i \rightarrow j}(\sigma_i) \mu_{j \rightarrow i}(\sigma_j), \quad (3.44)$$

$$\mu_{i, \partial i}(\sigma_i, \underline{\sigma}^{\partial i}) = \frac{1}{Z_{i, \partial i}} |\phi_i(\sigma_i)|^2 \prod_{j \in \partial i} |\phi_{ij}(\sigma_i, \sigma_j)|^2 \mu_{j \rightarrow i}(\sigma_j), \quad (3.45)$$

given in terms of the cavity marginals

$$\mu_{i \rightarrow j}(\sigma_i) \propto |\phi_i(\sigma_i)|^2 \prod_{k \in \partial i \setminus j} \left( \sum_{\sigma_k} |\phi_{ik}(\sigma_i, \sigma_k)|^2 \mu_{k \rightarrow i}(\sigma_k) \right). \quad (3.46)$$

For  $B_i = 0$  we can simplify the equations by taking the symmetric (or paramagnetic) solution of the BP equations. This is of course exact for a tree classical interaction graph  $\mathcal{E}_c$ . Then, the average local energies are given by

$$\langle e_i(\sigma_i, \underline{\sigma}^{\partial i}) \rangle_\mu = -h_i \prod_{j \in \partial i} \left( \frac{\cos(2K_{ij}^I)}{\cosh(2K_{ij}^R)} \right), \quad (3.47)$$

$$\langle e_{ij}(\sigma_i, \sigma_j) \rangle_\mu = -J_{ij} \tanh(2K_{ij}^R). \quad (3.48)$$

The resulting Min-Sum equations are

$$\mathcal{M}_{i \rightarrow j}(K_{ij}) = \min_{\{K_{ik} | k \in \partial i \setminus j\}} \left\{ \langle e_i \rangle_\mu + \sum_{k \in \partial i \setminus j} (\langle e_{ik} \rangle_\mu + \mathcal{M}_{k \rightarrow i}(K_{ik})) \right\}, \quad (3.49)$$

and the optimal couplings are estimated by

$$K_{ij}^0 = \arg \min_{K_{ij}} \{ \langle e_{ij} \rangle_\mu + \mathcal{M}_{i \rightarrow j}(K_{ij}) + \mathcal{M}_{j \rightarrow i}(K_{ij}) \}. \quad (3.50)$$

The orthogonality condition for two symmetric states reads

$$\langle \Psi_n(\underline{K}^n) | \Psi_m(\underline{K}^m) \rangle \propto \sum_{\underline{\sigma}} e^{\sum_{(ij) \in \mathcal{E}_c} (K_{ij}^{n*} + K_{ij}^m) \sigma_i \sigma_j} \propto \prod_{(ij) \in \mathcal{E}_c} \cosh(K_{ij}^{n*} + K_{ij}^m) = 0, \quad (3.51)$$

using the symmetric solution of the BP equations. Thus, to have orthogonality we need  $\cosh(K_{ij}^{n*} + K_{ij}^m) = 0$  for some link  $(ij)$ . That is, for the  $n$ th excited state we have the following set of constraints:

$$\mathbb{I}_{n,m} = \begin{cases} K_{i_m j_m}^{nR} + K_{i_m j_m}^{mR} = 0, \\ K_{i_m j_m}^{nI} - K_{i_m j_m}^{mI} = \frac{\pi}{2}, \end{cases} \quad (3.52)$$

for  $m = 0, \dots, n-1$ . The link  $(i_m j_m)$  is chosen to minimize the local energy  $\langle e_{i_m j_m} \rangle_{\mu_n}$ . Then, one can use the same Min-Sum equations as above to minimize over the remaining parameters. The number of orthogonal states that we can find in this way is limited by the number of the coupling parameters in the classical system which for a tree structure is  $N-1$ . The Hamiltonian matrix elements are given by

$$H_{mm'} = \prod_{(ij) \in \mathcal{E}_q} \left( \frac{\cosh(K_{ij}^{m*} + K_{ij}^{m'})}{(\cosh(2K_{ij}^{mR}) \cosh(2K_{ij}^{m'R}))^{1/2}} \right) \times \left\{ - \sum_{(ij) \in \mathcal{E}_q} J_{ij} \tanh(K_{ij}^{m*} + K_{ij}^{m'}) - \sum_i h_i \prod_{j \in \partial i} \left( \frac{\cosh(K_{ij}^{m*} - K_{ij}^{m'})}{\cosh(K_{ij}^{m*} + K_{ij}^{m'})} \right) \right\}. \quad (3.53)$$

Similarly, we can make more general orthogonal states  $|\Psi_n(\underline{B}^n, \underline{K}^n)\rangle$  and  $|\Psi_n(\underline{B}^m, \underline{K}^m)\rangle$  by choosing  $B_i^n$  such that for the complex measure  $\nu_{nm} \propto e^{\sum_i (B_i^{n*} + B_i^m) \sigma_i + \sum_{(ij) \in \mathcal{E}_c} (K_{ij}^{n*} + K_{ij}^m) \sigma_i \sigma_j}$

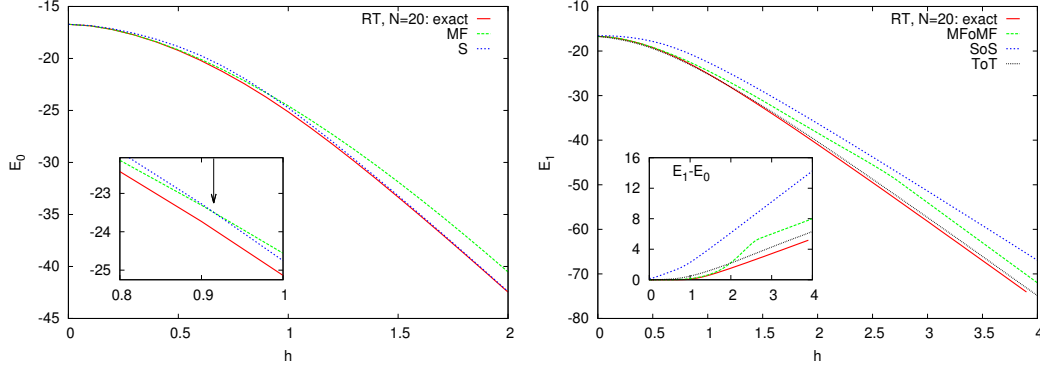


FIGURE 3.5: The ground-state and excited-state energies ( $E_0$  and  $E_1$ ) for the transverse Ising model on a random tree (RT) of size  $N = 20$  with random Gaussian couplings  $J_{ij}$  of mean zero and variance one in a uniform transverse field  $h_i = h$ . The exact results are compared with the upper bounds that are obtained by a minimal energy subspace spanned by product states (MFoMF), symmetric states (SoS), and tree states (ToT).

we have  $e^{\Delta F_i} = 0$  for some  $i$ . This gives

$$e^{2(B_i^{n*} + B_i^m)} = - \prod_{j \in \partial i} \left( \frac{\sum_{\sigma_j} e^{-(K_{ij}^{n*} + K_{ij}^m)\sigma_j} \nu_{j \rightarrow i}(\sigma_j)}{\sum_{\sigma_j} e^{(K_{ij}^{n*} + K_{ij}^m)\sigma_j} \nu_{j \rightarrow i}(\sigma_j)} \right). \quad (3.54)$$

The node  $i$  can be chosen in a greedy way to minimize  $\langle e_i(\sigma_i, \underline{\sigma}^{\partial i}) \rangle_{\mu_n}$  in the mean-field approximation. Here it is more difficult to minimize the average energy  $\langle E(\sigma) \rangle_{\mu_n}$ , which depends not only on the variational parameters but also on the BP cavity marginals  $\mu_{i \rightarrow j}(\sigma_i)$ . More precisely, the Min-Sum equations read

$$\mathcal{M}_{i \rightarrow j}(K_{ij}, \mu_{ij}) = \min_{B_i, \{K_{ik}, \mu_{ik} | k \in \partial i \setminus j\}: \mathbb{I}_{BP}^{(i)}} \left\{ \langle e_i \rangle_{\mu} + \sum_{k \in \partial i \setminus j} (\langle e_{ik} \rangle_{\mu} + \mathcal{M}_{k \rightarrow i}(K_{ik}, \mu_{ik})) \right\}, \quad (3.55)$$

where  $\mu_{ij} \equiv (\mu_{i \rightarrow j}, \mu_{j \rightarrow i})$ . Note that the minimum in the right hand side is conditioned on satisfying the local BP equations. The reader can find more details in Ref. [15].

### 3.3.4 Results

Let us start with a small system to compare the above approximations with the exact results for the ground and excited states. We take a random tree with random Gaussian couplings  $J_{ij}$  of mean zero and variance one in uniform transverse

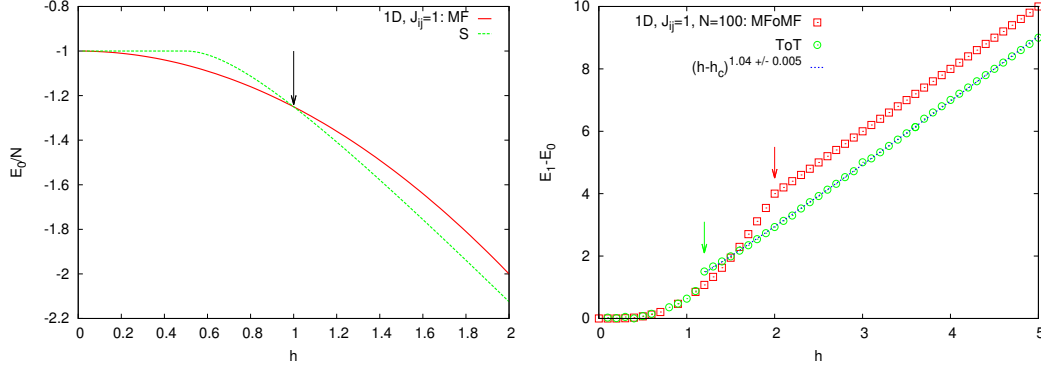


FIGURE 3.6: Left panel: The ground-state energy density ( $E_0/N$ ) for the one-dimensional (1D) transverse Ising model with uniform and ferromagnetic couplings ( $J_{ij} = 1$ ) obtained by the product states (MF) and the symmetric states (S) in the thermodynamic limit. Right panel: The energy gap  $E_1 - E_0$  obtained by the product states (MFoMF) and tree states (TOT) for the same model with  $N = 100$  spins. In both the cases the gap is non-analytic at the corresponding phase transition point, but it is non-vanishing close to the transition due to the local nature of the orthogonality constraints.

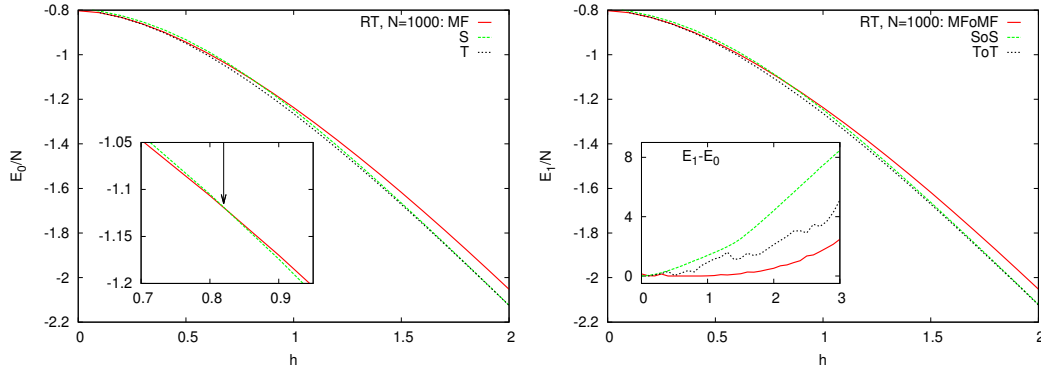


FIGURE 3.7: The ground-state and excited-state energy densities ( $E_0/N$  and  $E_1/N$ ) for the transverse Ising model on a random tree (RT) of size  $N = 1000$  with random Gaussian couplings  $J_{ij}$  of mean zero and variance one in a uniform transverse field  $h_i = h$ . We compare the upper bounds that are obtained by a minimal energy subspace spanned by product states (MFoMF), symmetric states (SoS), and tree states (ToT).

fields  $h_i = h$ . This system displays a phase transition from the ordered phase for  $h < h_c$  with a nonzero Edwards-Anderson order parameter  $q \equiv \sum_i \langle \sigma_i^z \rangle^2 / N$  to a disordered phase for  $h > h_c$ .

As Fig. 3.5 shows, we obtain better ground-state energies with the product and symmetric states in the ordered and disordered phases, respectively. While the product states allow for a nonzero magnetization, the symmetric states have by definition zero magnetization and therefore more appropriate to represent the disordered ground state. Indeed, an estimate of the transition point can be obtained

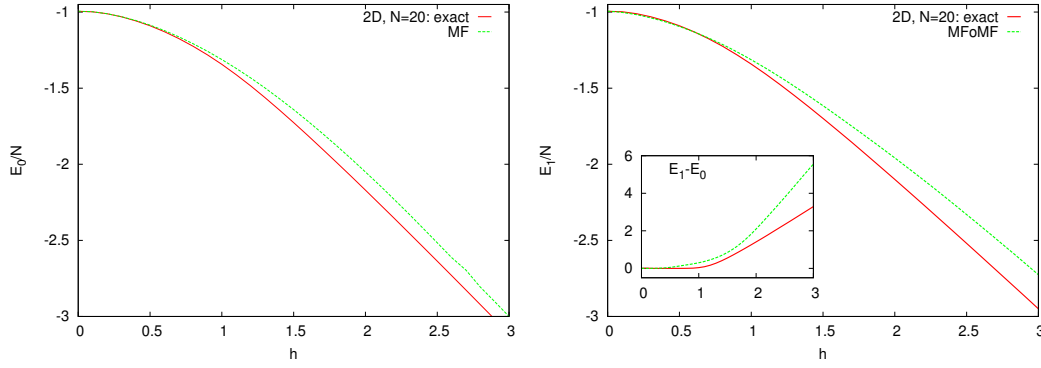


FIGURE 3.8: The ground-state and excited-state energy densities ( $E_0/N$  and  $E_1/N$ ) for the transverse Ising model on a two-dimensional (2D) square lattice of size  $N = 4 \times 5$  with random Gaussian couplings  $J_{ij}$  of mean zero and variance one in a uniform transverse field  $h_i = h$ . The exact results are compared with the upper bounds that are obtained by a minimal energy subspace spanned by product states (MFoMF).

by comparing the ground-state energies that are computed by the product and symmetric states. Figure 3.6 displays these energies for the one-dimensional transverse Ising model with ferromagnetic couplings. Nevertheless, for small system sizes we always obtain better energies for the excited state by a minimal energy subspace of the product states. In Fig. 3.5 we also display the results obtained by more general trial wave functions having a tree structure defined by  $\mathcal{E}_c = \mathcal{E}_q$ . As expected, we obtain much better upper bounds by introducing both the variational parameters  $B_i$  and  $K_{ij}$ . Figure 3.7 shows the results for a larger number of spins. Here, in the disordered phase we find better upper bounds for the excited-state energy by the symmetric trial wave function. This is due to the presence of very small couplings that reduce the cost of the orthogonality constraint.

Finally, we present the MF results for the transverse Ising model in a two-dimensional square lattice. In Fig. 3.8 we compare the upper bounds for the ground and excited states with the exact ones in a small lattice.

### 3.4 Discussion

In the section 3.3 it is shown a generalized variational quantum cavity method to study low-temperature excitations of quantum systems within the Bethe approximation. We constructed orthogonal sets of minimal energy quantum states, where the Hamiltonian matrix elements can be computed exactly to obtain some upper



bounds for the Hamiltonian eigenvalues. For more general trial wave functions we have only an approximate estimation of the Hamiltonian matrix elements but the estimation is expected to be asymptotically exact as long as the trial wave functions are represented by locally tree-like classical interaction graphs. And finally, the method can also be extended to include some appropriate global interactions in the trial wave functions that are essential to deal with the fermion sign problem [70].

# Chapter 4

## Variational Quantum Density Matrix

### 4.1 Introduction

This last chapter is divided in two parts. In the first one a new method to find approximate wave functions is described. The wave function is approximated in the same way presented in chapter 3, i.e. by a product of *local* terms. The algorithm proposed applies iteratively the imaginary time evolution operator to find the ground state wave function. This algorithm is able to find better solution and in more efficient way than the variational quantum cavity method presented in the previous chapter.

The imaginary time approach slowly modifies the parameters during the evolution in order to minimize the energy. This procedure permit to avoid the binning of the parameters used in the variational cavity method. This allows to use better approximate wave functions. The whole procedure is presented in the density matrix formalism because it is easier to extend it to finite temperature systems.

In the second part of the chapter we discuss a method to find approximate density matrices at finite temperature. In this case the density matrix is approximate by a product of *local* reduced density matrices. The algorithm performs an annealing in temperature, starting from infinite temperature. We force the density matrix, at each small temperature reduction, to remain a product of local reduced density matrices. The results obtained are encouraging, but the accumulation of the error

during the annealing process, due to the ansatz of the density matrix, reduces the precision of the algorithm.

Improved results can be obtained, from both algorithms presented in this chapter, using better expansions of the density matrix and working with generalized belief propagation equations [71]. This should be the case, above all, when we deal with square lattice problems [72].

In section 4.2 it is shown the imaginary time evolution of a Bethe density matrix. In sec. 4.3 it is described the finite temperature algorithms.

## 4.2 Imaginary time evolution

In this chapter I work directly with the density matrix formalism instead of the wave matrix formalism developed in chapter 3.

Let's consider an  $N$  body hamiltonian  $H$  with real parameters. Taking a base  $|\sigma\rangle$ , the density operator at zero temperature (isolated system)  $\rho$  is defined as:

$$\rho = |\psi\rangle \langle\psi| = \sum_{\sigma, \sigma'} C^*(\sigma) C(\sigma') |\sigma\rangle \langle\sigma'| \quad (4.1)$$

where we have expanded the wave function  $|\psi\rangle$  over a complete base  $\{|\sigma\rangle\}$  and  $C(\sigma) = \langle\psi|\sigma\rangle$ . If the wave function  $|\psi\rangle$  is normalized, the diagonal elements of the density matrix are probabilities, they define a measure. We proceed now in the same way of the variational quantum cavity method, see section 3.2. We have two different graphs, one  $\mathcal{E}_q$  generated by the physical interactions due to the hamiltonian, and another graph  $\mathcal{E}_c$  generated by the approximated measure. From now, I will consider  $\mathcal{E}_c = \mathcal{E}_q = \mathcal{E}$ . Let me approximate the probability distribution  $|C(\sigma)|^2$  with a Bethe form:

$$|C(\sigma)|^2 \sim \prod_i \psi_i^2(\sigma_i) \prod_{ij} \frac{\psi_{ij}^2(\sigma_i, \sigma_j)}{\psi_i^2(\sigma_i) \psi_j^2(\sigma_j)}. \quad (4.2)$$

We consider  $\psi_i$  and  $\psi_{ij}$  positive real functions, for simplicity. The  $\psi_i^2(\sigma_i) = \rho_i(\sigma_i)$  and  $\psi_{ij}^2(\sigma_i, \sigma_j) = \rho_{ij}(\sigma_i, \sigma_j)$  are respectively the one body and two body marginals

of the system. We define a zero temperature Bethe density matrix as:

$$\rho(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \prod_i \rho_i^{\frac{1}{2}}(\sigma_i) \rho_i^{\frac{1}{2}}(\sigma'_i) \prod_{ij} \frac{\rho_{ij}^{\frac{1}{2}}(\sigma_i, \sigma_j) \rho_{ij}^{\frac{1}{2}}(\sigma'_i, \sigma'_j)}{\rho_i^{\frac{1}{2}}(\sigma_i) \rho_i^{\frac{1}{2}}(\sigma'_i) \rho_j^{\frac{1}{2}}(\sigma_j) \rho_j^{\frac{1}{2}}(\sigma'_j)} \quad (4.3)$$

When the interaction graph of  $\mathcal{E}_c$  generated by a zero temperature Bethe density matrix is a tree, it is easy to show that the following identities hold<sup>1</sup>:

$$\begin{aligned} \rho_i(\sigma_i) &= \sum_{\boldsymbol{\sigma} \setminus \sigma_i} \langle \boldsymbol{\sigma} | \rho | \boldsymbol{\sigma} \rangle = \sum_{\boldsymbol{\sigma} \setminus \sigma_i} \rho(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \\ \rho_{ij}(\sigma_i, \sigma_j) &= \sum_{\boldsymbol{\sigma} \setminus \sigma_i \sigma_j} \langle \boldsymbol{\sigma} | \rho | \boldsymbol{\sigma} \rangle = \sum_{\boldsymbol{\sigma} \setminus \sigma_i \sigma_j} \rho(\boldsymbol{\sigma}, \boldsymbol{\sigma}). \end{aligned} \quad (4.4)$$

More accurate zero temperature density matrices can be obtained by considering interactions between a larger number of variables, for example,

$$\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_{(ij) \in \mathcal{E}_c} \rho_{ij}(\sigma_i, \sigma_j)^{\frac{1}{2}} \rho_{ij}(\sigma'_i, \sigma'_j)^{\frac{1}{2}} \prod_i \frac{\rho_{i\partial i}(\sigma_i, \sigma_{\partial i})^{\frac{1}{2}} \rho_{i\partial i}(\sigma'_i, \sigma'_{\partial i})^{\frac{1}{2}}}{\prod_{k \in \partial i} \rho_{ik}(\sigma_i, \sigma_k)^{\frac{1}{2}} \rho_{ik}(\sigma'_i, \sigma'_k)^{\frac{1}{2}}}, \quad (4.5)$$

where  $\sigma_{\partial i} = \{\sigma_j | j \in \partial i\}$ , and  $\partial i$  denotes the neighborhood set of  $i$  in the interaction graph  $\mathcal{E}_c$ . The same identity equations (4.4) hold for the measure (4.5) in the case of tree interaction graph  $\mathcal{E}_c$ . We have also the following one:

$$\rho_{i\partial i}(\sigma_i, \sigma_{\partial i}) = \sum_{\boldsymbol{\sigma} \setminus \sigma_i, \sigma_{\partial i}} \langle \boldsymbol{\sigma} | \rho | \boldsymbol{\sigma} \rangle = \sum_{\boldsymbol{\sigma} \setminus \sigma_i, \sigma_{\partial i}} \rho(\boldsymbol{\sigma}, \boldsymbol{\sigma}). \quad (4.6)$$

### 4.2.1 Imaginary time evolutions of a Bethe density matrix.

The imaginary time evolution is a technique that, in principle, allows to compute the ground state of quantum systems. The idea is to apply to a trial density matrix the imaginary time evolution operator  $e^{-tH}$ . The excited states are exponentially suppressed in the limit of  $t \rightarrow \infty$ . Let's consider a density matrix  $\rho$  and let's apply the imaginary time operator to it:

$$e^{-tH} \rho e^{-tH} = \sum_{n,m} e^{-tH} C_m^* C_n |E_n\rangle \langle E_m| e^{-tH} = \sum_{n,m} e^{-t(E_n+E_m)} C_m^* C_n |E_n\rangle \langle E_m| \quad (4.7)$$

---

<sup>1</sup>This identity can be derived considering the normalizations and marginalization constraints over the marginals  $\rho_i(\sigma_i)$  and  $\rho_{ij}(\sigma_i, \sigma_j)$ , and starting the trace over the variables from leafs.

where we have expanded over a base where the hamiltonian is diagonal. Taking the limit of  $t \rightarrow \infty$  we obtain:

$$\lim_{t \rightarrow \infty} e^{-tH} \rho e^{-tH} \approx e^{-2tE_g} |C_g|^2 |E_g\rangle \langle E_g| \quad (4.8)$$

where the  $|E_g\rangle$  is the ground state of the system. If we apply, iteratively, the operator  $e^{-\epsilon H}$  to a trial density matrix and if the initial trial density matrix has an overlap different from zero with the ground state, the final outcome, at least approximatively, is the ground state.

We want to apply iteratively the imaginary time evolution operator  $e^{-\epsilon H}$  in order to find an approximated ground state matrix. Let's consider, for simplicity, a zero temperature Bethe density matrix at imaginary time  $t$ :

$$\rho^t(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \prod_i \sqrt{\rho_i^t(\sigma_i) \rho_i^t(\sigma'_i)} \prod_{ij} \frac{\sqrt{\rho_{ij}^t(\sigma_i, \sigma_j) \rho_{ij}^t(\sigma'_i, \sigma'_j)}}{\sqrt{\rho_i^t(\sigma_i) \rho_i^t(\sigma'_i) \rho_j^t(\sigma_j) \rho_j^t(\sigma'_j)}} \quad (4.9)$$

the density matrix at time  $t + \epsilon$  reads:

$$\rho^{t+\epsilon} = \frac{e^{-\epsilon H} \rho^t e^{-\epsilon H}}{\text{Tr}\{e^{-\epsilon H} \rho^t e^{-\epsilon H}\}} \quad (4.10)$$

where I normalized the expression (4.7). Now we assume that also the density matrix  $\rho^{t+\epsilon}$  has a Bethe form and so we compute the quantities  $\rho_i^{t+\epsilon}$  and  $\rho_{ij}^{t+\epsilon}$  through the formula (4.4). We need to compute the marginal probabilities  $\rho_i(\sigma_i)$  and  $\rho_{ij}(\sigma_i, \sigma_j)$  of the following system:

$$\text{Tr}\{\rho^{t+\epsilon}\} = \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | \rho^{t+\epsilon} | \boldsymbol{\sigma} \rangle \quad (4.11)$$

$$\propto \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | e^{-\epsilon H} \rho^t e^{-\epsilon H} | \boldsymbol{\sigma} \rangle \quad (4.12)$$

$$= \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | e^{-2\epsilon H} \rho^t | \boldsymbol{\sigma} \rangle \quad (4.13)$$

$$= \sum_{\boldsymbol{\sigma}, \boldsymbol{\sigma}'} \langle \boldsymbol{\sigma} | e^{-2\epsilon H} | \boldsymbol{\sigma}' \rangle \langle \boldsymbol{\sigma}' | \rho^t | \boldsymbol{\sigma} \rangle \quad (4.14)$$

where in the third line I used the cyclic property of trace, and in the fourth I inserted the identity operator  $\mathbb{1} = \sum_{\boldsymbol{\sigma}'} |\boldsymbol{\sigma}'\rangle \langle \boldsymbol{\sigma}'|$ . The marginals can be computed in the limit of  $\epsilon \ll 1$ . In effect, let's consider a quantum interaction graph  $\mathcal{E}_q$  generated by an hamiltonian  $H$ , the  $\langle \boldsymbol{\sigma} | e^{-\epsilon H} | \boldsymbol{\sigma}' \rangle$  can be decomposed, by a Suzuki-Trotter

expansion [73, 74], in a product of functions that have the same graph structure of  $\mathcal{E}_q$ . Then if the hamiltonian  $H$  has local interactions, also  $\langle \sigma | e^{-\epsilon H} | \sigma' \rangle$  is composed by products of local functions [75]. The eq.(4.11) is a product of local quantities and the marginal of this system can be computed by the BP equations. In the next section this procedure is applied to the transverse Ising model.

### 4.2.2 Tranverse ising model

Consider the transverse field Ising model with Hamiltonian  $H = \sum_{(ij) \in \mathcal{E}_q} H_{ij} + \sum_i H_i$  where  $H_{ij} \equiv -J_{ij} \sigma_i^z \sigma_j^z$ ,  $H_i \equiv -h_i \sigma_i^x$ , and  $i = 1, \dots, N$  labels the sites in the quantum interaction graph  $\mathcal{E}_q$ . The  $\sigma_i^{x,y,z}$  are the standard Pauli matrices. We assume that  $h_i \geq 0$ . In the following we will work in the  $\sigma^z$  representation with orthonormal basis  $|\sigma\rangle \equiv |\sigma_1 \sigma_2 \dots \sigma_N\rangle$ . The Suzuki-Trotter [73, 74] expansion allows to simplify the exponential of two non commutative quantities. In the case of the exponential of the transverse Ising hamiltonian we can write:

$$e^{-\epsilon H} = \prod_{(ij) \in \mathcal{E}_q} e^{\epsilon J_{ij} \sigma_i^z \sigma_j^z / 2} \prod_i e^{\epsilon h_i \sigma_i^x} \prod_{(ij) \in \mathcal{E}_q} e^{\epsilon J_{ij} \sigma_i^z \sigma_j^z / 2} + O(\epsilon^3). \quad (4.15)$$

The matrix element  $\langle \sigma | e^{-\epsilon H} | \sigma' \rangle$  with the approximation above is:

$$\langle \sigma | e^{-\epsilon H} | \sigma' \rangle \approx \prod_i w_i(\sigma_i, \sigma'_i) \prod_{(ij) \in \mathcal{E}_q} w_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j) \quad (4.16)$$

where

$$w_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j) \equiv e^{\epsilon J_{ij} (\sigma_i \sigma_j + \sigma'_i \sigma'_j) / 2}, \quad (4.17)$$

$$w_i(\sigma_i, \sigma'_i) \equiv \cosh(\epsilon h_i) \delta_{\sigma_i, \sigma'_i} + \sinh(\epsilon h_i) \delta_{\sigma_i, -\sigma'_i}. \quad (4.18)$$

Taking a Bethe zero temperature density matrix (4.9) at imaginary time  $t$  we have to find marginals at time  $t + \epsilon$  of the following probability distribution:

$$\begin{aligned} \rho^{\epsilon+t}(\sigma) &= \sum_{\sigma'} \langle \sigma | e^{-2\epsilon H} | \sigma' \rangle \langle \sigma' | \rho^t | \sigma \rangle = \\ &= \sum_{\sigma'} \prod_i w_i(\sigma_i, \sigma'_i) \sqrt{\rho_i^t(\sigma_i) \rho_i^t(\sigma'_i)} \prod_{(ij) \in \mathcal{E}_q} w_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j) \frac{\sqrt{\rho_{ij}^t(\sigma_i, \sigma_j) \rho_{ij}^t(\sigma'_i, \sigma'_j)}}{\sqrt{\rho_i^t(\sigma_i) \rho_i^t(\sigma'_i) \rho_j^t(\sigma_j) \rho_j^t(\sigma'_j)}}. \end{aligned} \quad (4.19)$$

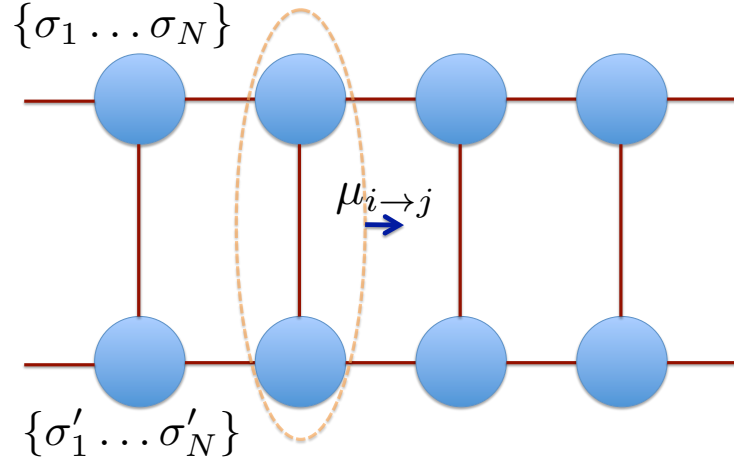


FIGURE 4.1: In this figure it's shown a 1 dimension two layers system. The messages  $\mu_{i \rightarrow j}$  group the two layers variables.

It is a two layers classic system, see fig. 4.1. Then, the BP equations over the system are:

$$\mu_{i \rightarrow j}^{n+1}(\sigma_j \sigma'_j) = \sum_{\sigma_i \sigma'_i} w_i w_{ij} \frac{\sqrt{\rho_{ij}^t \rho_{ij}^{t'}}}{\sqrt{\rho_j^t \rho_j^{t'}}} \prod_{k \in \partial i \setminus j} \mu_{k \rightarrow i}^n(\sigma_i \sigma'_i) \quad (4.20)$$

where the index  $n$  refers to the iteration of BP equations, and  $\rho_{ij}^{t'} \equiv \rho_{ij}^t(\sigma'_i, \sigma'_j)$ . Given the set of fixed point messages  $\{\mu_{i \rightarrow j}^*\}$  of the eq.(4.20), the marginals can be computed by the following formulas:

$$\rho_i^{t+\epsilon}(\sigma_i) = \sum_{\sigma'_i} w_i \sqrt{\rho_j^t \rho_j^{t'}} \prod_{j \in \partial i} \mu_{j \rightarrow i}^*(\sigma_i \sigma'_i) \quad (4.21)$$

$$\rho_{ij}^{t+\epsilon}(\sigma_i, \sigma_j) = \sum_{\sigma'_i, \sigma'_j} w_{ij} \sqrt{\rho_{ij}^t \rho_{ij}^{t'}} \prod_{p \in \partial i \setminus j} \mu_{p \rightarrow i}^*(\sigma_i \sigma'_i) \prod_{q \in \partial j \setminus i} \mu_{q \rightarrow j}^*(\sigma_j \sigma'_j). \quad (4.22)$$

We can summarize the procedure of finding the zero temperature Bethe density matrix by the following steps:

1. Initialize the marginals  $\rho_i^0$  and  $\rho_{ij}^0$  with a flat or random distribution.
2. Run the BP equations (4.20) until convergence.
3. Update the marginals, eq.(4.21).
4. Return to point 2 until the marginals converge.

In general we are looking for the fixed point solution of the set of marginals  $\{\rho_i, \rho_{ij}$  and messages  $\mu_{i \rightarrow j}$  of the equations (4.21) and (4.20). The procedure above can be simplified when we consider homogenous systems where we assume that each one-body, two-body marginals and the messages are the same in the whole system. We can consider the fixed point equations for both marginals and BP equations at the same time. The problem of finding the zero temperature Bethe density matrix reduces to find the quantities  $\rho_i$ ,  $\rho_{ij}$  and  $\mu$  that solves the following set of equations:

$$\begin{cases} \mu(\sigma_j \sigma'_j) = \sum_{\sigma_i \sigma'_i} w_i w_{ij} \frac{\sqrt{\rho_{ij} \rho'_{ij}}}{\sqrt{\rho_j \rho'_j}} \mu^{d-1}(\sigma_i \sigma'_i) \\ \rho_i(\sigma_i) = \sum_{\sigma'_i} w_i \sqrt{\rho_j \rho'_j} \mu^d(\sigma_i \sigma'_i) \\ \rho_{ij}(\sigma_i, \sigma_j) = \sum_{\sigma'_i, \sigma'_j} w_{ij} \sqrt{\rho_{ij} \rho'_{ij}} \mu^{d-1}(\sigma_i \sigma'_i) \mu^{d-1}(\sigma_j \sigma'_j) \end{cases} \quad (4.23)$$

where  $d$  is the number of neighbors of one node. In the next subsection I will show numerical results obtained by c++ programs that implement the procedure described above for different approximated measures: mean-field, Bethe measure eq.(4.3) and near neighbors region eq.(4.5). The programs can take as input whatever transverse Ising model hamiltonian. The only restriction is due to the convergence problems of BP equations over particular interaction graphs. Moreover when we deal with interaction graphs with loops, the average quantities computed are Bethe approximations of the real ones.

### 4.2.3 Numerical results

The first question that we face is: is this procedure able to really minimize the approximate wave function? In figure 4.2 are shown the energies of a transverse Ising chain computed by a zero temperature pairwise density matrix eq.(4.3). These energies are compared with those found by the variational quantum cavity method with the same structure of the wave function. As we can see from the plot 4.2, there is no difference between the two energies.

The graphs 4.4 and 4.3 are the magnetizations respectively on chain and on random regular graph of degree three computed by imaginary time evolution. The blue curve is the magnetization computed by mean-field approximation, the red one is the pairwise density matrix eq.(4.3), and the green one is computed by neighbors



region eq.(4.5). The exact transition point is located at  $h_c = 1$  and  $h_c = 2.232$  [13] respectively for chain and for random regular graph with degree three. The critical transition fields found by the neighbors region zero temperature density matrix are closer to the correct ones than those found by pairwise and mean-field approximations (see captions of the figures 4.4 and 4.3 for precise values of the critical fields).

Moreover we run the algorithm on small disorder system in order to compare the energies on loopy graphs to the exact one found by brute force diagonalization of the hamiltonian. We consider the gaps between the exact energies  $E_{ex}$  and those found by imaginary time evolution with pairwise  $E_{pw}$  and neighbor region  $E_{nr}$  density matrix approximations. The gap is defined as  $100 * (E_{ex} - E_{approx})/E_{ex}$ . In the fig 4.5 we show the gaps computed between the exact energies and those found by imaginary time evolution of two different matrix approximations on a 2d lattice with 20 spins (4 times 5 spins) with periodic boundary conditions. In this case the  $J_{ij} = 1$  and the transverse fields  $h_i$  are random numbers uniformly distributed in  $[0, h]$ . In the same plot the magnetizations  $M^z$  is shown (dashed lines). We can note that the gaps are always under 0.5. They differ from zero only near the quantum critical point where the magnetization in the  $z$  direction drops to zero. Also in this case using larger regions improve the results.

### 4.3 Annealing of quantum systems

In this section we face the problem to find approximate density matrices at finite temperature  $T$ .

The equilibrium density matrix at inverse finite temperature  $\beta$  should satisfy the following equation [76]:

$$\frac{\partial \rho(\beta)}{\partial \beta} = -H\rho \quad \text{with} \quad \rho(0) = \mathbb{1} \quad (4.24)$$

where  $H$  is the hamiltonian of the system. Then giving the density matrix at inverse temperature  $\beta$ ,  $\rho^\beta$ , we may write:

$$\rho^{\beta+\epsilon} = \frac{e^{-\epsilon H} \rho^\beta}{Z^{\beta+\epsilon}} \quad (4.25)$$

## Energy density of chain

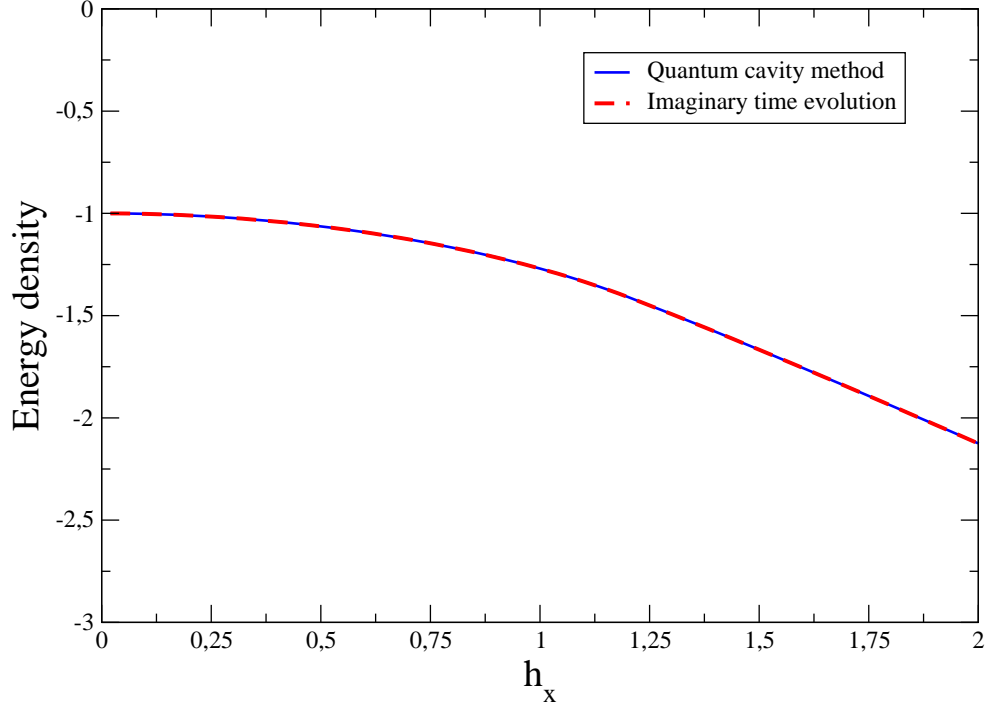


FIGURE 4.2: In this figure it is shown the energy density found on chain with  $J_{ij} = 1$  with two different methods that minimize a trial density matrix in the form eq.(4.3). The blue line is computed by the quantum cavity method presented in the chapter 3 and the red one by imaginary time evolution. There is no significant difference between them. The imaginary time step used is  $\epsilon = 10^{-3}$

where  $Z^{\beta+\epsilon} = \text{Tr}\{e^{-\epsilon H} \rho^\beta\}$  is a normalization constant. We want to give an approximate description of the density matrix that is locally and globally consistent on tree structures. The idea is to write the density matrix as a product of reduced density matrices. In the mean-field approximation the density matrix is simply approximated by  $\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_i \rho_i(\sigma_i; \sigma'_i)$ . In the Bethe approximation we may write the density matrix as

$$\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_i \rho_i(\sigma_i; \sigma'_i) \prod_{(ij) \in \mathcal{E}} \frac{\rho_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j)}{\rho_i(\sigma_i; \sigma'_i) \rho_j(\sigma_j; \sigma'_j)}, \quad (4.26)$$

for a locally tree-like interaction graph  $\mathcal{E}_c$  that could be different from the quantum interaction graph  $\mathcal{E}_q$ . Here we consider  $\mathcal{E}_c = \mathcal{E}_q = \mathcal{E}$ . Note that this is an ansatz for the matrix elements of  $\rho$  that is Hermitian but not necessarily semi-positive. In appendix B we see that when the interaction graph  $\mathcal{E}$  is a tree and the reduced

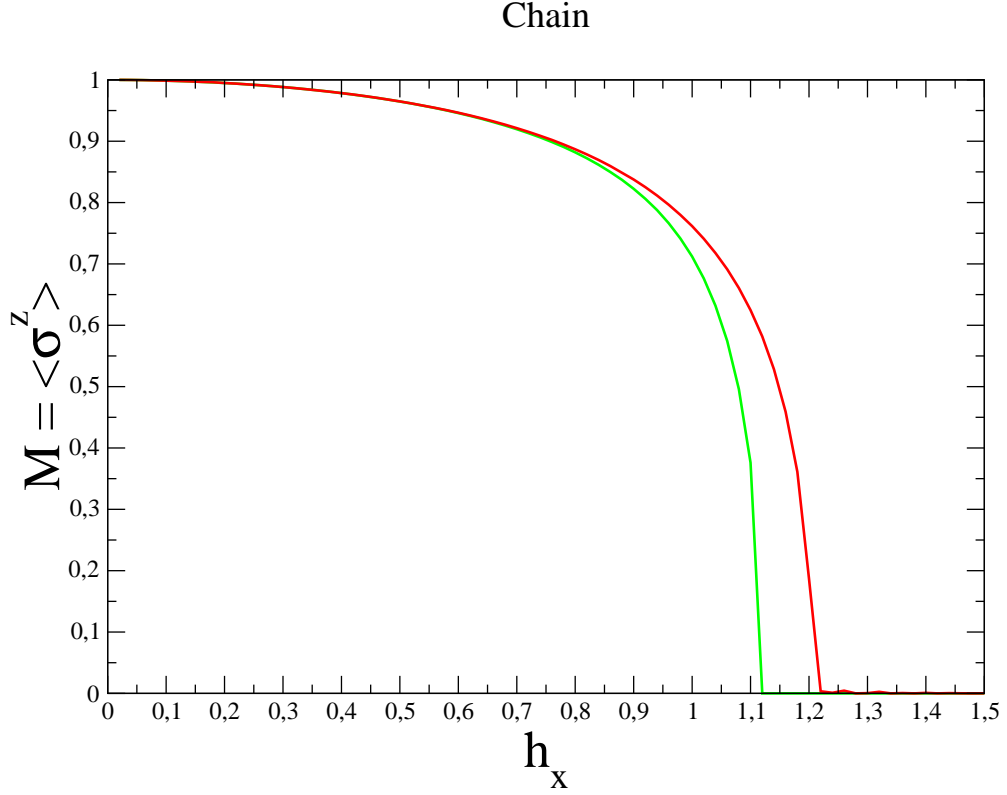


FIGURE 4.3: The plot shows the magnetization  $M = \langle \sigma^z \rangle$  of a chain with  $J_{ij} = 1$  and varying  $h_x$ . The number of spins are  $N = 50$  with periodic boundary conditions. The imaginary time step used is  $\epsilon = 10^{-3}$ . The red line is the Bethe zero temperature density matrix (pw) eq.(4.3). The green one is the neighbors region zero temperature density matrix (nn) eq.(4.5). The critical fields found are respectively  $h_c^{pw} = 1.22$  and  $h_c^{nn} = 1.12$ . The exact one is  $h_c = 1$ .

density matrices  $\rho_{ij}$  and  $\rho_i$  are locally consistent we have  $\rho_{ij} = \text{Tr}_{\setminus i,j} \rho$  and  $\rho_i = \text{Tr}_{\setminus i} \rho$ .

More accurate density matrices can be obtained by considering interactions between a larger number of variables, for example,

$$\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_{(ij) \in \mathcal{E}} \rho_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j) \prod_i \frac{\rho_{i\partial i}(\sigma_i, \sigma_{\partial i}; \sigma'_i, \sigma'_{\partial i})}{\prod_{k \in \partial i} \rho_{ik}(\sigma_i, \sigma_k; \sigma'_i, \sigma'_k)}, \quad (4.27)$$

where  $\sigma_{\partial i} = \{\sigma_j | j \in \partial i\}$ , and  $\partial i$  denotes the neighborhood set of  $i$  in the interaction graph  $\mathcal{E}$ . In the same lines of appendix B we can show that for tree interaction graphs  $\mathcal{E}$  and locally consistent  $\rho_{i\partial i}$  and  $\rho_{ij}$  we have  $\rho_{i\partial i} = \text{Tr}_{\setminus i, \partial i} \rho$  and  $\rho_{ij} = \text{Tr}_{\setminus i,j} \rho$ .

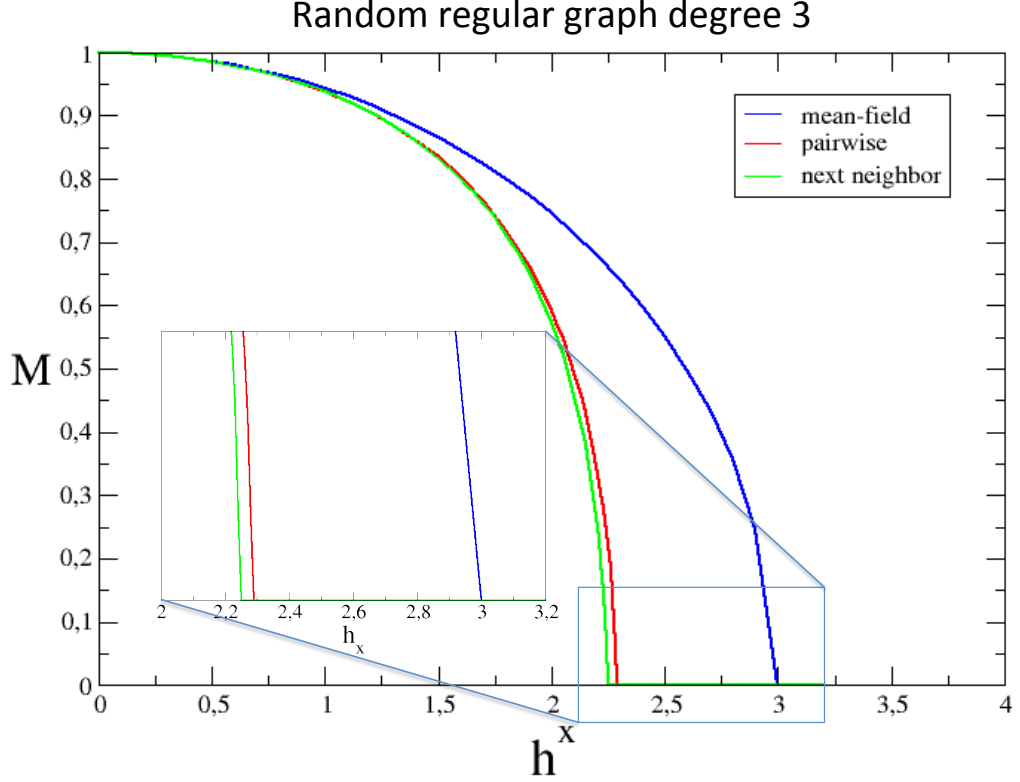


FIGURE 4.4: The plot shows the magnetization  $M = \langle \sigma^z \rangle$  of a random regular graph with degree three with  $J_{ij} = 1$  and varying  $h_x$ . The number of spins are  $N = 50$ . The imaginary time step used is  $\epsilon = 10^{-3}$ . The blue line is the mean-field zero temperature density matrix, the red one is the Bethe zero temperature density matrix eq.(4.3) and the green one is the neighbors region zero temperature density matrix eq.(4.5). The critical fields found are respectively  $h_c^{mf} = 3$ ,  $h_c^{pw} = 2.29$  and  $h_c^{nn} = 2.25$ . The exact one is  $h_c = 2.232$ .

### 4.3.1 Annealing algorithm of quantum Ising model

In order to find the density matrix describing the equilibrium state of the system at temperature  $T = 1/\beta$  we start from the density matrix at infinite temperature  $\rho \propto \mathbb{I}$  and we decrease slowly the temperature in an annealing process. We use a similar procedure described for the imaginary time evolution in subsection 4.2.2. The density matrix  $\rho^{\beta+\epsilon}$  reads:

$$\rho^{\beta+\epsilon}(\boldsymbol{\sigma}; \boldsymbol{\sigma}') \propto \sum_{\boldsymbol{\sigma}''} \prod_i w_i(\sigma_i, \sigma_i'') \prod_{(ij) \in \mathcal{E}_q} w_{ij}(\sigma_i, \sigma_j; \sigma_i'', \sigma_j'') \rho^\beta(\boldsymbol{\sigma}''; \boldsymbol{\sigma}'), \quad (4.28)$$

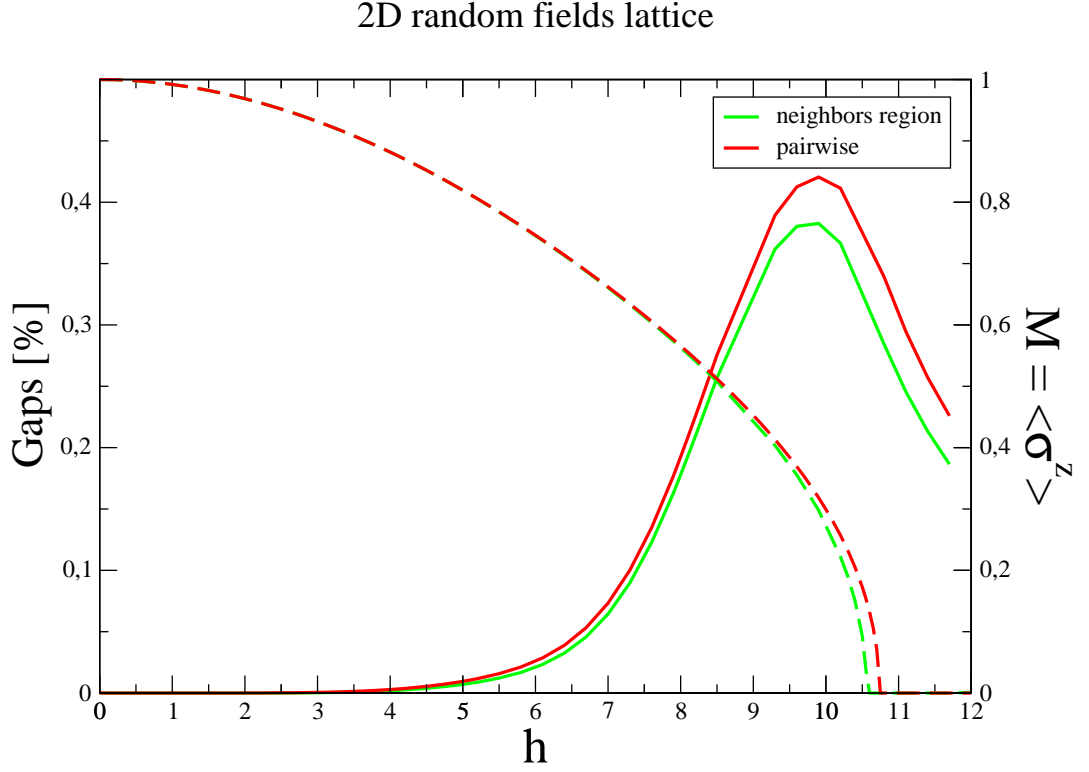


FIGURE 4.5: The plot shows the gaps  $(100 * (E_{ex} - E_{approx})/E_{ex})$  between the exact energies and two different approximate energies found by imaginary time evolution. The red one is founded using a Bethe density matrix eq. (4.3), and the green one by neighbors region density matrix eq. (4.5). Dashed lines show the magnetizations  $M^z = \langle \sigma^z \rangle$  of the same approximate density matrices. We note that the gaps are different from zero only near the quantum transition point, where the magnetizations drop to zero. The gaps are always below 0.5. Better results are obtained with neighbors region approximate density matrix.

where

$$w_{ij}(\sigma_i, \sigma_j; \sigma_i'', \sigma_j'') \equiv e^{\epsilon J_{ij}(\sigma_i \sigma_j + \sigma_i'' \sigma_j'')/2}, \quad (4.29)$$

$$w_i(\sigma_i, \sigma_i'') \equiv \cosh(\epsilon h_i) \delta_{\sigma_i, \sigma_i''} + \sinh(\epsilon h_i) \delta_{\sigma_i, -\sigma_i''}. \quad (4.30)$$

Let us start from the mean-field approximation of the density matrix  $\rho = \prod_i \rho_i$  where in each step the density matrix is updated as follows

$$\rho^{\beta+\epsilon}(\boldsymbol{\sigma}; \boldsymbol{\sigma}') \propto \sum_{\boldsymbol{\sigma}''} \prod_i \left[ w_i(\sigma_i, \sigma_i'') \rho_i^\beta(\sigma_i'', \sigma_i') \right] \prod_{(ij) \in \mathcal{E}_q} w_{ij}(\sigma_i, \sigma_j; \sigma_i'', \sigma_j''). \quad (4.31)$$

Then, within the Bethe approximation the reduced density matrices  $\rho_i^{\beta+\epsilon} = \text{Tr}_{\setminus i} \rho^{\beta+\epsilon}$  are obtained by

$$\rho_i^{\beta+\epsilon}(\sigma_i; \sigma'_i) \propto \sum_{\sigma''_i} \left[ w_i(\sigma_i, \sigma''_i) \rho_i^\beta(\sigma''_i; \sigma'_i) \right] \prod_{j \in \partial i} \left( \sum_{\sigma''_j} w_{ij}(\sigma_i, \sigma_j; \sigma''_i, \sigma''_j) \mu_{j \rightarrow i}(\sigma_j, \sigma''_j) \right). \quad (4.32)$$

Here  $\partial i$  denotes the neighborhood set of  $i$  in  $\mathcal{E}_q$ , and the cavity marginals  $\mu_{i \rightarrow j}(\sigma_i, \sigma''_i)$  are determined by the BP equations for the Gibbs measure  $\rho^{\beta+\epsilon}(\boldsymbol{\sigma}; \boldsymbol{\sigma})$ ,

$$\mu_{i \rightarrow j}(\sigma_i, \sigma''_i) \propto \sum_{\sigma''_i} \left[ w_i(\sigma_i, \sigma''_i) \rho_i^\beta(\sigma''_i; \sigma_i) \right] \prod_{k \in \partial i \setminus j} \left( \sum_{\sigma''_k} w_{ik}(\sigma_i, \sigma_k; \sigma''_i, \sigma''_k) \mu_{k \rightarrow i}(\sigma_k, \sigma''_k) \right). \quad (4.33)$$

Given the  $\rho_i$  and the weights  $w_i, w_{ij}$ , we solve the above equations by iteration starting from random initial cavity marginals  $\mu_{i \rightarrow j}(\sigma_i, \sigma''_i)$ , and use the cavity marginals to find the lower-temperature reduced density matrices  $\rho_i^{\beta+\epsilon}$ . The equations showed in this section is very close to those presented in sec. 4.2 for the imaginary time evolution algorithm. Indeed the imaginary time procedure, started from a flat distribution, can be seen as an annealing process where at each step we force the density matrix to have a zero temperature structure like in the equation (4.3). The main difference between the two procedures relies in the form of the density matrix: the diagonal elements are parametrized in a different way.

### 4.3.2 Numerical result

The figure 4.6 shows the phase transition points that are obtained in this way for the ferromagnetic transverse Ising model on a random regular graph. Unfortunately, the errors in each step of the annealing algorithm accumulate, giving rise to larger and larger errors as we decrease the temperature. The point is that at each step of the annealing process we assume that the present density matrix  $\rho$  is the right density matrix at inverse temperature  $\beta$  which is correct only if we work with the most general density matrix representation. As a result the density matrix that we obtain is not the optimal one; indeed, minimizing the free energy directly at inverse temperature  $\beta$  with the same density matrix representation could result to smaller free energies [14]. However, as the figure shows, the error is reduced by enlarging the space of the trial density matrices. In appendix C we give the update equations for neighbor region density matrices.

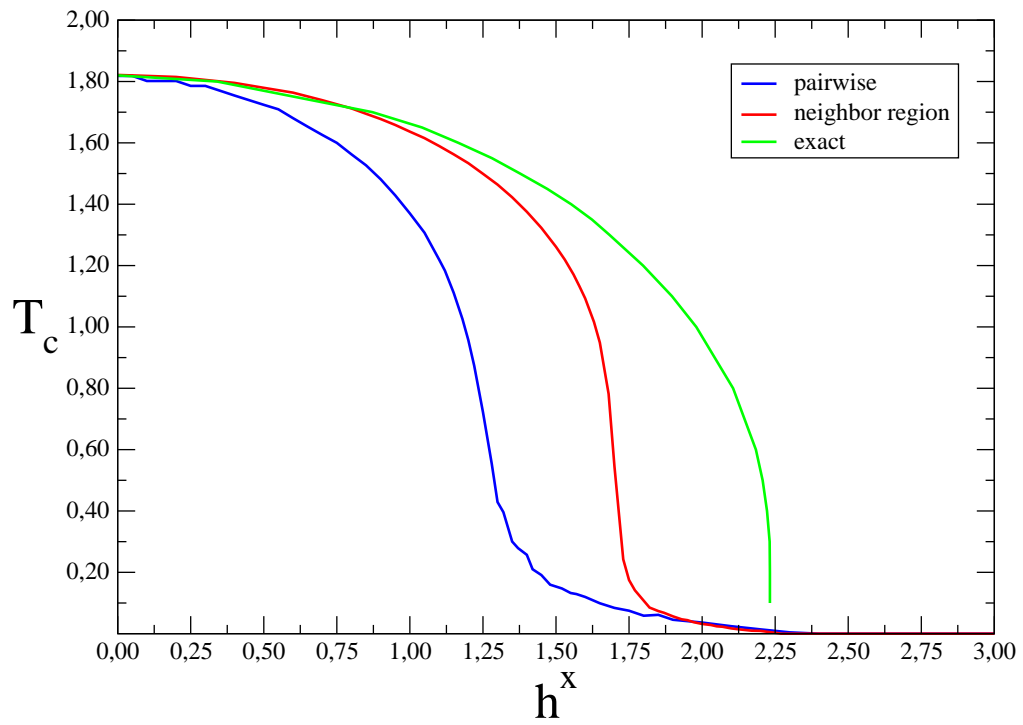


FIGURE 4.6: Comparing the phase transition points in the transverse Ising model ( $J_{ij} = 1$  and  $h_i = h$ ) on a random regular graph of degree  $K = 3$  obtained by the annealing algorithm (for system size  $N = 50$  and different density matrix representations) and the exact solution of the path integral quantum cavity method [13] in the thermodynamic limit.

# Appendix A

## Post-processing and optimality proofs

Before tackling the proof of the Theorem 2, we will need the following definitions and a technical result.

**Definition 7.** (Computation tree) The computation tree is a cover of the graph  $G$ , in the following sense: it is an (infinite) tree  $\mathcal{T}_G$  along with an application  $\pi : \mathcal{T}_G \rightarrow G$  that satisfies (a)  $\pi$  is surjective and (b)  $\pi|_{i \cup \partial i} : i \cup \partial i \rightarrow \pi(i \cup \partial i)$  is a graph isomorphism for every  $i \in \mathcal{T}_G$ . It can be explicitly constructed as the graph of non-backtracking paths in  $G$  starting on a given node  $v_0$ , with two paths being connected iff the longest one is identical to the other except for an additional final node (and edge). Up to graph isomorphisms, this tree does not depend on the choice  $v_0$ .

The (finite) tree  $\mathcal{T}_G(t, v_0)$  is defined by the radius  $t$  ball centered around  $v_0$  in  $\mathcal{T}_G$ . Alternatively, it can be directly constructed as the graph of non-backtracking paths of length up to  $t$  starting on  $v_0$ , with two paths being connected iff the longest one is identical to the other except for an additional final node (and edge). Clearly the finite computation tree depends strongly on the choice of  $v_0$ .

For both computation trees, edge weights (and node prizes) will be lifted (transported) naturally as  $c_{ij} = c_{\pi(i)\pi(j)}$ .

Lifting edge constraints by  $g_{ij} = g_{\pi(i)\pi(j)}$  defines a  $(R, D)$ -PCSF problem with  $R = \pi^{-1}(\{r\})$  on  $\mathcal{T}_G$ . On  $\mathcal{T}_G(t, v_0)$  instead, it gives a slightly relaxed  $(R, D)$ -PCSF



problem in which leaf nodes can point to neighbors in  $G$  that are not present in  $\mathcal{T}_G$ . For convenience, let us extend  $\pi$  by setting  $\pi(*) = *$ .

*Remark 8.* As  $\mathcal{T}_G(t, v)$  is a tree, the Max-Sum equations are exact and have a unique fixed point in  $\mathcal{T}_G(t, v)$  [27].

**Lemma 9.** *Any Max-Sum fixed point in a graph  $G$  can be naturally lifted to a Max-Sum fixed point in  $\mathcal{T}_G$ . Moreover, any Max-Sum fixed point can be naturally lifted to a Max-Sum fixed point over a slightly modified  $\mathcal{T}_G(t, v)$  with extra cost terms only on leaves.*

*Proof.* As Max-Sum equations are local and the two graphs are locally isomorphic, given a fixed point  $\{\psi_{ij}\}_{(i,j) \in E}$ , the messages  $\Psi_{ij} = \psi_{\pi(i)\pi(j)}$  satisfy the fixed point equations on  $\mathcal{T}_G$ . On  $\mathcal{T}_G(t, v)$  the Max-Sum equations are satisfied everywhere except possibly on leaf nodes (where the graphs are not locally isomorphic). Given a leaf  $i$  attached with edge  $(i, j)$ , add an energy term  $-E_i(d_i p_i) = \psi_{\pi(i)\pi(j)}(d_i, \pi(p_i))$ . Now Max-Sum equations are satisfied everywhere on for this modified cost function.  $\square$

Now we proceed to prove Theorem 2

### A.0.3 Proof of Theorem 2

*Proof.* Assume  $S'$  oriented towards the root node  $r$ , i.e. defining a parenthood vector  $(p'_i)_{i \in V'}$ , such that  $E' = \{(i, p'_i) : i \in V' \setminus \{r\}\}$ . Consider the subgraph  $S = (V_S, E_S)$  of  $\mathcal{T}_G(N+1, r)$  induced by  $S^*$ , i.e. defined by  $V_S = \{v : \pi(v) \in V^*\}$ ,  $E_S = \{(i, j) : (\pi(i), \pi(j)) \in E^*\}$ .

It can be easily proven that the connected component in  $S$  of the root node of  $\mathcal{T}_G(N+1, r)$  is a tree  $S''$  isomorphic to  $S^*$  (see [51]). Denote by  $\{p^*\}$  the decisional variables induced by  $S^*$  and by  $\{p'\}$  the ones induced by  $S'$ . The parenthood assignment

$$q_i = \begin{cases} p'_i & i \in V_{S''} \\ p_i^* & i \notin V_{S''} \end{cases}$$

satisfies  $q_i \neq *$  if  $q_j = i$  (as  $V' \subseteq V^*$ ) and so depths  $d_i$  can be assigned so as to verify all  $g_{ij}$  constraints in  $\mathcal{T}_G(N+1, r)$ . Now the cost associated with  $\mathbf{q}$  is  $\mathcal{H}(\mathbf{q}) = \sum_{i \in V_{S''}} c_{ip'_i} + \sum_{i \notin V_{S''}} c_{ip_i^*} \geq \sum_{i \in \mathcal{T}_G(N+1, r)} c_{ip_i^*} = \sum_{i \in V_{S''}} c_{ip_i^*} + \sum_{i \notin V_{S''}} c_{ip_i^*}$

due to the optimality of the MS solution  $\mathbf{p}^*$  in the computation tree (this is because MS is always exact on a tree). This implies clearly that  $\mathcal{H}(S^*) \leq \mathcal{H}(S')$ .  $\square$

# Appendix B

## Locally and globally consistent reduced density matrices

Consider the following Bethe ansatz for the density matrix

$$\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_i \rho_i(\sigma_i; \sigma'_i) \prod_{(ij) \in \mathcal{E}} \frac{\rho_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j)}{\rho_i(\sigma_i; \sigma'_i) \rho_j(\sigma_j; \sigma'_j)}. \quad (\text{B.1})$$

Here we prove that when  $\mathcal{E}$  is a tree and the reduced density matrices are locally consistent we have  $\rho_{ij} = \text{Tr}_{\setminus i,j} \rho$  and  $\rho_i = \text{Tr}_{\setminus i} \rho$ .

Let us start from computing  $\text{Tr} \rho$  to show that  $\rho$  is trace normalized when  $\rho_i = \text{Tr}_j \rho_{ij}$  and  $\text{Tr} \rho_i = \text{Tr} \rho_{ij} = 1$ . Expanding the trace we have

$$Z = \text{Tr} \rho = \sum_{\boldsymbol{\sigma}} \prod_i \rho_i(\sigma_i; \sigma_i) \prod_{(ij) \in \mathcal{E}} \frac{\rho_{ij}(\sigma_i, \sigma_j; \sigma_i, \sigma_j)}{\rho_i(\sigma_i; \sigma_i) \rho_j(\sigma_j; \sigma_j)}. \quad (\text{B.2})$$

For tree structures we can compute the above sum recursively

$$Z = \sum_{\sigma_i} \rho_i(\sigma_i; \sigma_i) \prod_{j \in \partial i} \left( \sum_{\sigma_j} \frac{\rho_{ij}(\sigma_i, \sigma_j; \sigma_i, \sigma_j)}{\rho_i(\sigma_i; \sigma_i) \rho_j(\sigma_j; \sigma_j)} Z_{j \rightarrow i}(\sigma_j) \right), \quad (\text{B.3})$$

using the Bethe equations [27],

$$Z_{i \rightarrow j}(\sigma_i) = \rho_i(\sigma_i; \sigma_i) \prod_{k \in \partial i \setminus j} \left( \sum_{\sigma_k} \frac{\rho_{ik}(\sigma_i, \sigma_k; \sigma_i, \sigma_k)}{\rho_i(\sigma_i; \sigma_i) \rho_k(\sigma_k; \sigma_k)} Z_{k \rightarrow i}(\sigma_k) \right), \quad (\text{B.4})$$

where  $Z_{i \rightarrow j}(\sigma_i)$  is the cavity partition function when spin  $i$  has state  $\sigma_i$ . Note that for the leaves we have  $Z_{i \rightarrow j}(\sigma_i) = \rho_i(\sigma_i; \sigma_i)$  and from the marginalization relations  $\rho_i = \text{Tr}_j \rho_{ij}$  we find  $Z_{i \rightarrow j}(\sigma_i) = \rho_i(\sigma_i; \sigma_i)$  for all the cavity partition functions. Therefore we obtain  $Z = \sum_{\sigma_i} \rho_i(\sigma_i; \sigma_i) = 1$ .

To compute the one-body reduced density matrices we use again the recursive equations to write

$$\langle \sigma_i | \text{Tr}_{\setminus i} \rho | \sigma'_i \rangle = \rho_i(\sigma_i; \sigma'_i) \prod_{j \in \partial i} \left( \sum_{\sigma_j} \frac{\rho_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma_j)}{\rho_i(\sigma_i; \sigma'_i) \rho_j(\sigma_j; \sigma_j)} Z_{j \rightarrow i}(\sigma_j) \right). \quad (\text{B.5})$$

But  $Z_{j \rightarrow i}(\sigma_j) = \rho_j(\sigma_j; \sigma_j)$  which along with the marginalization relations give  $\text{Tr}_{\setminus i} \rho = \rho_i$ . Similarly, one can prove that  $\text{Tr}_{\setminus i, j} \rho = \rho_{ij}$  thanks to the tree interaction graph  $\mathcal{E}$  and the consistency of the local density matrices  $\rho_i$  and  $\rho_{ij}$ .

One can easily extend the above arguments to more general density matrices with higher order interactions

$$\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_i \rho_i(\sigma_i; \sigma'_i) \prod_a \frac{\rho_a(\sigma_{\partial a}; \sigma'_{\partial a})}{\prod_{i \in \partial a} \rho_i(\sigma_i; \sigma'_i)}, \quad (\text{B.6})$$

as long as the bipartite graph representing the dependency of the interactions to the variables is a tree. Here  $\sigma_{\partial a} \equiv \{\sigma_i | i \in \partial a\}$  and  $\partial a$  defines the set of variables in interaction  $a$ .

# Appendix C

## Computing the reduced density matrices in the annealing algorithm

Consider the following Bethe ansatz for the density matrix

$$\rho(\boldsymbol{\sigma}; \boldsymbol{\sigma}') = \prod_i \rho_i(\sigma_i; \sigma'_i) \prod_{(ij) \in \mathcal{E}} \frac{\rho_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j)}{\rho_i(\sigma_i; \sigma'_i) \rho_j(\sigma_j; \sigma'_j)}. \quad (\text{C.1})$$

In each step of the annealing process we need to compute the local reduced density matrices given the updated density matrix

$$\tilde{\rho}(\boldsymbol{\sigma}; \boldsymbol{\sigma}') \propto \sum_{\boldsymbol{\sigma}''} \prod_i \left[ w_i(\sigma_i, \sigma''_i) \rho_i(\sigma''_i; \sigma'_i) \right] \prod_{(ij) \in \mathcal{E}} \left[ w_{ij}(\sigma_i, \sigma_j; \sigma''_i, \sigma''_j) \frac{\rho_{ij}(\sigma''_i, \sigma''_j; \sigma'_i, \sigma'_j)}{\rho_i(\sigma''_i; \sigma'_i) \rho_j(\sigma''_j; \sigma'_j)} \right]. \quad (\text{C.2})$$

The local density matrix  $\tilde{\rho}_{ij} = \text{Tr}_{\setminus i,j} \tilde{\rho}$  is obtained from the above expression after summing over the  $\sigma_k = \sigma'_k$  for  $k \neq i, j$ . For tree interaction graphs  $\mathcal{E}$  this sum can be computed by considering the cavity messages that the boundary variables

$\partial(ij)$  receive from the other parts of the system in addition to the local weights,

$$\begin{aligned}
\tilde{\rho}_{ij}(\sigma_i, \sigma_j; \sigma'_i, \sigma'_j) &\propto \sum_{\sigma''_i, \sigma''_j} w_{ij}(\sigma_i, \sigma_j; \sigma''_i, \sigma''_j) \rho_{ij}(\sigma''_i, \sigma''_j; \sigma'_i, \sigma'_j) \\
&\times w_i(\sigma_i, \sigma''_i) \prod_{k \in \partial i \setminus j} \left[ \sum_{\sigma_k, \sigma''_k} w_{ik}(\sigma_i, \sigma_k; \sigma''_i, \sigma''_k) \frac{\rho_{ik}(\sigma''_i, \sigma''_k; \sigma'_i, \sigma_k)}{\rho_i(\sigma''_i; \sigma'_i) \rho_k(\sigma''_k; \sigma_k)} \mu_{k \rightarrow i}(\sigma_k; \sigma''_k) \right] \\
&\times w_j(\sigma_j, \sigma''_j) \prod_{k \in \partial j \setminus i} \left[ \sum_{\sigma_k, \sigma''_k} w_{jk}(\sigma_j, \sigma_k; \sigma''_j, \sigma''_k) \frac{\rho_{jk}(\sigma''_j, \sigma''_k; \sigma'_j, \sigma_k)}{\rho_j(\sigma''_j; \sigma'_j) \rho_k(\sigma''_k; \sigma_k)} \mu_{k \rightarrow j}(\sigma_k; \sigma''_k) \right].
\end{aligned} \tag{C.3}$$

Here the cavity messages  $\mu_{i \rightarrow j}$  are determined recursively by the Bethe equations,

$$\begin{aligned}
\mu_{i \rightarrow j}(\sigma_i; \sigma''_i) &\propto w_i(\sigma_i; \sigma''_i) \rho_i(\sigma''_i; \sigma_i) \\
&\times \prod_{k \in \partial i \setminus j} \left[ \sum_{\sigma_k, \sigma''_k} w_{ik}(\sigma_i, \sigma_k; \sigma''_i, \sigma''_k) \frac{\rho_{ik}(\sigma''_i, \sigma''_k; \sigma_i, \sigma_k)}{\rho_i(\sigma''_i; \sigma_i) \rho_k(\sigma''_k; \sigma_k)} \mu_{k \rightarrow i}(\sigma_k; \sigma''_k) \right].
\end{aligned} \tag{C.4}$$

# Bibliography

- [1] P. W. Anderson. More is different. *Science*, 177(4047):393–396, August 1972. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.177.4047.393. URL <http://www.sciencemag.org/content/177/4047/393>. PMID: 17796623.
- [2] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin Glass Theory and Beyond*. World Scientific Publishing Company, September 1987. ISBN 9971501163.
- [3] M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, August 2002. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1073287. URL <http://www.sciencemag.org/content/297/5582/812>. PMID: 12089451.
- [4] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms*, 27(2): 201–226, 2005. ISSN 1098-2418. doi: 10.1002/rsa.20057. URL <http://onlinelibrary.wiley.com/doi/10.1002/rsa.20057/abstract>.
- [5] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina. Polynomial iterative algorithms for coloring and analyzing random graphs. *Physical Review E*, 68(3):036702, September 2003. doi: 10.1103/PhysRevE.68.036702. URL <http://link.aps.org/doi/10.1103/PhysRevE.68.036702>.
- [6] Christos H Papadimitriou. *Computational complexity*. Addison-Wesley, Reading, Mass., 1994. ISBN 0201530821 9780201530827 0020153082 9780020153085.
- [7] Christos H Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Prentice Hall, Englewood Cliffs, N.J., 1982. ISBN 0131524623 9780131524620 0486402584 9780486402581.

- [8] Indaco Biazzo, Alfredo Braunstein, and Riccardo Zecchina. Performance of a cavity-method-based algorithm for the prize-collecting steiner tree problem on graphs. *Physical Review E*, 86(2):026706, August 2012. doi: 10.1103/PhysRevE.86.026706. URL <http://link.aps.org/doi/10.1103/PhysRevE.86.026706>.
- [9] GB Dantzig. Maximization of a linear function of variables subject to linear inequalities, in activity analysis of production and allocation. Wiley, 1951.
- [10] L. De Cesare, K. Lukierska Walasek, and K. Walasek. Cavity-field approach to quantum spin glasses: The ising spin glass in a transverse field. *Physical Review B*, 45(14):8127–8130, April 1992. doi: 10.1103/PhysRevB.45.8127. URL <http://link.aps.org/doi/10.1103/PhysRevB.45.8127>.
- [11] M.S. Leifer and D. Poulin. Quantum graphical models and belief propagation. *Annals of Physics*, 323(8):1899–1946, August 2008. ISSN 0003-4916. doi: 10.1016/j.aop.2007.10.001. URL <http://www.sciencedirect.com/science/article/pii/S0003491607001509>.
- [12] C. Laumann, A. Scardicchio, and S. L. Sondhi. Cavity method for quantum spin glasses on the bethe lattice. *Physical Review B*, 78(13):134424, October 2008. doi: 10.1103/PhysRevB.78.134424. URL <http://link.aps.org/doi/10.1103/PhysRevB.78.134424>.
- [13] Florent Krzakala, Alberto Rosso, Guilhem Semerjian, and Francesco Zamponi. Path-integral representation for quantum spin models: Application to the quantum cavity method and monte carlo simulations. *Physical Review B (Condensed Matter and Materials Physics)*, 78(13):134428, 2008. doi: 10.1103/PhysRevB.78.134428. URL <http://link.aps.org/abstract/PRB/v78/e134428>.
- [14] L. B. Ioffe and Marc Mézard. Disorder-driven quantum phase transitions in superconductors and magnets. *Physical Review Letters*, 105(3):037001, July 2010. doi: 10.1103/PhysRevLett.105.037001. URL <http://link.aps.org/doi/10.1103/PhysRevLett.105.037001>.
- [15] A. Ramezanpour. Cavity approach to variational quantum mechanics. *Physical Review B*, 85(12):125131, March 2012. doi: 10.1103/PhysRevB.85.125131. URL <http://link.aps.org/doi/10.1103/PhysRevB.85.125131>.



- [16] F. Altarelli, A. Braunstein, A. Ramezanpour, and R. Zecchina. Stochastic matching problem. *Physical Review Letters*, 106(19):190601, May 2011. doi: 10.1103/PhysRevLett.106.190601. URL <http://link.aps.org/doi/10.1103/PhysRevLett.106.190601>.
- [17] I. Biazzo and A. Ramezanpour. Low-temperature excitations within the bethe approximation. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(04):P04011, April 2013. ISSN 1742-5468. doi: 10.1088/1742-5468/2013/04/P04011. URL <http://iopscience.iop.org/1742-5468/2013/04/P04011>.
- [18] Marc Mézard, G. Parisi, and M. A. Virasoro. SK model: The replica solution without replicas. *EPL (Europhysics Letters)*, 1(2):77, January 1986. ISSN 0295-5075. doi: 10.1209/0295-5075/1/2/006. URL <http://iopscience.iop.org/0295-5075/1/2/006>.
- [19] M. Mézard and G. Parisi. The bethe lattice spin glass revisited. *The European Physical Journal B - Condensed Matter and Complex Systems*, 20(2):217–233, March 2001. ISSN 1434-6028, 1434-6036. doi: 10.1007/PL00011099. URL <http://link.springer.com/article/10.1007/PL00011099>.
- [20] H. A. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, 150(871):552–575, July 1935. ISSN 1364-5021, 1471-2946. doi: 10.1098/rspa.1935.0122. URL <http://rspa.royalsocietypublishing.org/content/150/871/552>.
- [21] Judea Pearl. Reverend bayes on inference engines: a distributed hierarchical approach. In *in Proceedings of the National Conference on Artificial Intelligence*, page 133–136, 1982.
- [22] Robert G. Gallager. *Low-Density Parity-Check Codes*. 1963.
- [23] D. J. Thouless, P. W. Anderson, and R. G. Palmer. Solution of 'Solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, March 1977. ISSN 0031-8086. doi: 10.1080/14786437708235992. URL <http://adsabs.harvard.edu/abs/1977PMag...35..593T>.
- [24] Giorgio Parisi. On local equilibrium equations for clustering states. *eprint arXiv:cs/0212047*, page 12047, December 2002. URL <http://adsabs.harvard.edu/abs/2002cs.....12047P>.

- [25] Alfredo Braunstein and Riccardo Zecchina. Survey propagation as local equilibrium equations. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(06):P06007, June 2004. ISSN 1742-5468. doi: 10.1088/1742-5468/2004/06/P06007. URL <http://iopscience.iop.org/1742-5468/2004/06/P06007>.
- [26] F. Altarelli, A. Braunstein, A. Ramezanzpour, and R. Zecchina. Stochastic optimization by message passing. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(11):P11009, November 2011. ISSN 1742-5468. doi: 10.1088/1742-5468/2011/11/P11009. URL <http://iopscience.iop.org/1742-5468/2011/11/P11009>.
- [27] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, USA, March 2009. ISBN 9780198570837.
- [28] Francesco Zamponi. Mean field theory of spin glasses. arXiv e-print 1008.4844, August 2010. URL <http://arxiv.org/abs/1008.4844>.
- [29] Giovanni Gallavotti. *Statistical Mechanics: A Short Treatise*. Springer, August 1999. ISBN 9783540648833.
- [30] Giorgio Parisi. *Statistical Field Theory*. Perseus Books, 1998. ISBN 9780738200514.
- [31] Christos H. Papadimitriou. Computational complexity. In *Encyclopedia of Computer Science*, page 260–265. John Wiley and Sons Ltd., Chichester, UK, 2003. ISBN 0-470-86412-5. URL <http://dl.acm.org/citation.cfm?id=1074100.1074233>.
- [32] Ludwig Boltzmann. *Vorlesungen über Gastheorie: Th. Theorie des Gase mit einatomigen Molekülen, deren dimensionen gegen die Mittlere weglänge verschwinden*. J. A. Barth, 1896.
- [33] Claude Elwood Shannon. *The Mathematical Theory of Communication*. University of Illinois Press, 1949. ISBN 9780252725487.
- [34] Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, September 1986. ISSN 0004-3702. doi: 10.1016/0004-3702(86)90072-X. URL <http://www.sciencedirect.com/science/article/pii/000437028690072X>.

- [35] Rémi Monasson, Riccardo Zecchina, Scott Kirkpatrick, Bart Selman, and Lidror Troyansky. Determining computational complexity from characteristic ‘phase transitions’. *Nature*, 400(6740):133–137, July 1999. ISSN 0028-0836. doi: 10.1038/22055. URL <http://www.nature.com/nature/journal/v400/n6740/abs/400133a0.html>.
- [36] Olivier C. Martin, Rémi Monasson, and Riccardo Zecchina. Statistical mechanics methods and phase transitions in optimization problems. *Theoretical Computer Science*, 265(1–2):3–67, August 2001. ISSN 0304-3975. doi: 10.1016/S0304-3975(01)00149-9. URL <http://www.sciencedirect.com/science/article/pii/S0304397501001499>.
- [37] Andrea Montanari and Rudiger Urbanke. Modern coding theory: The statistical mechanics and computer science point of view. arXiv e-print 0704.2857, April 2007. URL <http://arxiv.org/abs/0704.2857>.
- [38] Marc Mézard. Optimization and physics: On the satisfiability of random boolean formulae. In Daniel Iagolnitzer, Vincent Rivasseau, and Jean Zinn-Justin, editors, *International Conference on Theoretical Physics*, pages 475–488. Birkhäuser Basel, January 2004. ISBN 978-3-0348-9618-4, 978-3-0348-7907-1. URL [http://link.springer.com/chapter/10.1007/978-3-0348-7907-1\\_37](http://link.springer.com/chapter/10.1007/978-3-0348-7907-1_37).
- [39] Leonid A Levin. Average case complete problems. *SIAM J. Comput.*, 15(1):285–286, February 1986. ISSN 0097-5397. doi: 10.1137/0215020. URL <http://dx.doi.org/10.1137/0215020>.
- [40] Yuri Gurevich. Average case completeness. *Journal of Computer and System Sciences*, 42(3):346–398, June 1991. ISSN 0022-0000. doi: 10.1016/0022-0000(91)90007-R. URL <http://www.sciencedirect.com/science/article/pii/002200009190007R>.
- [41] Shai Ben-david, Benny Chor, Oded Goldreich, and Michael Luby. On the theory of average case complexity. *Journal of Computer and System Sciences*, 44:193–219, 1997.
- [42] Shao-Shan Carol Huang and Ernest Fraenkel. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Science signaling*, 2(81):ra40, 2009. ISSN 1937-9145. doi: 10.1126/scisignal.2000350. PMID: 19638617.

- [43] M. Bailly-Bechet, A. Braunstein, and R. Zecchina. A prize-collecting steiner tree approach for transduction network inference. In *Computational Methods in Systems Biology*, page 83–95, 2009.
- [44] M. Bailly-Bechet, C. Borgs, A. Braunstein, J. Chayes, A. Dagkessamanskaia, J.-M. Francois, and R. Zecchina. Finding undetected protein associations in cell signaling by belief propagation. *Proceedings of the National Academy of Science*, 108:882–887, January 2011. ISSN 0027-8424. doi: 10.1073/pnas.1004751108;. URL <http://adsabs.harvard.edu/abs/2011PNAS..108..882B>.
- [45] J. Hackner. : *Energiewirtschaftlich optimale Ausbauplanung kommunaler Fernwärmesysteme*. PhD thesis, Vienna University of Technology, Austria, 2004.
- [46] Michel X. Goemans and David P. Williamson. The primal-dual method for approximation algorithms and its application to network design problems. In Dorit S. Hochbaum, editor, *Approximation algorithms for NP-hard problems*, page 144–191. PWS Publishing Co., Boston, MA, USA, 1997. ISBN 0-534-94968-1. URL <http://dl.acm.org/citation.cfm?id=241938.241942>.
- [47] David S Johnson, Maria Minkoff, and Steven Phillips. The prize collecting steiner tree problem: theory and practice. In *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, SODA '00, page 760–769, Philadelphia, PA, USA, 2000. Society for Industrial and Applied Mathematics. ISBN 0-89871-453-2. URL <http://portal.acm.org/citation.cfm?id=338219.338637>. ACM ID: 338637.
- [48] I. Ljubic, R. Weiskircher, U. Pferschy, G. Klau, P. Mutzel, and M. Fischetti. Solving the prize-collecting steiner tree problem to optimality. In *Proc. of the Seventh Workshop on Algorithm Engineering and Experiments (ALENEX 05)*, 2004.
- [49] Abilio Lucena and Mauricio G.C Resende. Strong lower bounds for the prize collecting steiner problem in graphs. *Discrete Applied Mathematics*, 141 (1–3):277–294, May 2004. ISSN 0166-218X. doi: 10.1016/S0166-218X(03)00380-9. URL <http://www.sciencedirect.com/science/article/pii/S0166218X03003809>.

- [50] M. Bayati, C. Borgs, A. Braunstein, J. Chayes, A. Ramezanpour, and R. Zecchina. Statistical mechanics of steiner trees. *Physical Review Letters*, 101(3):037208, July 2008. doi: 10.1103/PhysRevLett.101.037208. URL <http://link.aps.org/doi/10.1103/PhysRevLett.101.037208>.
- [51] Mohsen Bayati, Alfredo Braunstein, and Riccardo Zecchina. A rigorous analysis of the cavity equations for the minimum spanning tree. *Journal of Mathematical Physics*, 49(12):125206, December 2008. ISSN 00222488. doi: 10.1063/1.2982805. URL [http://jmp.aip.org/resource/1/jmapaq/v49/i12/p125206\\_s1?isAuthorized=no](http://jmp.aip.org/resource/1/jmapaq/v49/i12/p125206_s1?isAuthorized=no).
- [52] Alfredo Braunstein and Riccardo Zecchina. Learning by message passing in networks of discrete synapses. *Physical Review Letters*, 96(3):030201, January 2006. doi: 10.1103/PhysRevLett.96.030201. URL <http://link.aps.org/doi/10.1103/PhysRevLett.96.030201>.
- [53] Riccardo Zecchina and Alfredo Braunstein. pcmp, 2013. URL <http://areeweb.polito.it/ricerca/cmp/>.
- [54] Alexandre Salles da Cunha, Abilio Lucena, Nelson Maculan, and Mauricio G.C. Resende. A relax-and-cut algorithm for the prize-collecting steiner problem in graphs. *Discrete Applied Mathematics*, 157(6): 1198–1217, March 2009. ISSN 0166-218X. doi: 10.1016/j.dam.2008.02.014. URL <http://www.sciencedirect.com/science/article/B6TYW-4S9FGYT-1/2/b801553099711e9ee9c1d65c5d0ff9fc>.
- [55] K. Aardal and C. P. M. van Hoesel. Polyhedral techniques in combinatorial optimization i: Theory. *Statistica Neerlandica*, 50(1): 3–26, 1996. ISSN 1467-9574. doi: 10.1111/j.1467-9574.1996.tb01478.x. URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9574.1996.tb01478.x/abstract>.
- [56] George B. Dantzig and Mukund N. Thapa. *Linear Programming: 1: Introduction*. Springer, January 1997. ISBN 9780387948331.
- [57] Ivana Ljubic. Prize collecting steiner tree, 2013. URL <http://homepage.univie.ac.at/ivana.ljubic/research/pcstp/>.
- [58] Thorsten Koch. steinlib, 2013. URL <http://steinlib.zib.de/steinlib.php>.

- [59] Isabel Rosseti, Marcus Poggi de Aragão, Celso C. Ribeiro, Eduardo Uchoa, and Renato F. Werneck. New benchmark instances for the steiner problem in graphs. In *Metaheuristics: Computer Decision-Making*, number 86 in Applied Optimization, pages 601–614. Springer US, January 2004. ISBN 978-1-4419-5403-9, 978-1-4757-4137-7. URL [http://link.springer.com/chapter/10.1007/978-1-4757-4137-7\\_28](http://link.springer.com/chapter/10.1007/978-1-4757-4137-7_28).
- [60] Cees Duin and Stefan Voß. Efficient path and vertex exchange in steiner tree algorithms. *Networks*, 29(2):89–105, March 1997. ISSN 1097-0037. doi: 10.1002/(SICI)1097-0037(199703)29:2<89::AID-NET3>3.0.CO;2-7. URL [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0037\(199703\)29:2<89::AID-NET3>3.0.CO;2-7/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0037(199703)29:2<89::AID-NET3>3.0.CO;2-7/abstract).
- [61] H. Q. Lin, J. E. Gubernatis, Harvey Gould, and Jan Tobochnik. Exact diagonalization methods for quantum systems. *Computers in Physics*, 7(4):400–407, July 1993. ISSN 0894-1866. doi: 10.1063/1.4823192. URL <http://scitation.aip.org/content/aip/journal/cip/7/4/10.1063/1.4823192>.
- [62] N. Nakatsuji. Cluster expansion of the wavefunction. excited states. *Chemical Physics Letters*, 59(2):362–364, November 1978. ISSN 0009-2614. doi: 10.1016/0009-2614(78)89113-1. URL <http://www.sciencedirect.com/science/article/pii/0009261478891131>.
- [63] A. K. Theophilou. The energy density functional formalism for excited states. *Journal of Physics C: Solid State Physics*, 12(24):5419, December 1979. ISSN 0022-3719. doi: 10.1088/0022-3719/12/24/013. URL <http://iopscience.iop.org/0022-3719/12/24/013>.
- [64] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33–83, January 2001. doi: 10.1103/RevModPhys.73.33. URL <http://link.aps.org/doi/10.1103/RevModPhys.73.33>.
- [65] Victor Bapst and Guilhem Semerjian. On quantum mean-field models and their quantum annealing. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(06):P06007, June 2012. ISSN 1742-5468. doi: 10.1088/1742-5468/2012/06/P06007. URL <http://iopscience.iop.org/1742-5468/2012/06/P06007>.

- [66] Eduardo R. Gagliano, Elbio Dagotto, Adriana Moreo, and Francisco C. Alcaraz. Correlation functions of the antiferromagnetic heisenberg model using a modified lanczos method. *Physical Review B*, 34(3):1677–1682, August 1986. doi: 10.1103/PhysRevB.34.1677. URL <http://link.aps.org/doi/10.1103/PhysRevB.34.1677>.
- [67] Richard Courant and David Hilbert. *Methods of mathematical physics. Vol. 1 Vol. 1*. Wiley, New York, 1989. ISBN 0585294283 9780585294285. URL <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=26283>.
- [68] F.R. Kschischang, B.J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001. ISSN 0018-9448. doi: 10.1109/18.910572.
- [69] Robert Jastrow. Many-body problem with strong forces. *Physical Review*, 98(5):1479–1484, June 1955. doi: 10.1103/PhysRev.98.1479. URL <http://link.aps.org/doi/10.1103/PhysRev.98.1479>.
- [70] A. Ramezanpour and R. Zecchina. Sign problem in the bethe approximation. *Physical Review B*, 86(15):155147, October 2012. doi: 10.1103/PhysRevB.86.155147. URL <http://link.aps.org/doi/10.1103/PhysRevB.86.155147>.
- [71] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Generalized belief propagation. In *IN NIPS 13*, page 689–695. MIT Press, 2000.
- [72] Alessandro Pelizzola. Cluster variation method in statistical physics and probabilistic graphical models. *Journal of Physics A: Mathematical and General*, 38(33):R309, August 2005. ISSN 0305-4470. doi: 10.1088/0305-4470/38/33/R01. URL <http://iopscience.iop.org/0305-4470/38/33/R01>.
- [73] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959. ISSN 0002-9939, 1088-6826. doi: 10.1090/S0002-9939-1959-0108732-6. URL <http://www.ams.org/proc/1959-10-04/S0002-9939-1959-0108732-6/>.
- [74] Masuo Suzuki. Generalized trotter’s formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. *Communications in Mathematical Physics*, 51(2):183–190, June 1976. ISSN 0010-3616, 1432-0916. doi: 10.1007/BF01609348. URL <http://link.springer.com/article/10.1007/BF01609348>.

- 
- [75] Jean Zinn-Justin. *Path Integrals in Quantum Mechanics*. Oxford University Press, July 2010. ISBN 9780198566755.
- [76] Richard P. Feynman. *Statistical Mechanics: A Set Of Lectures*. Westview Press, 2 edition, March 1998. ISBN 0201360764.